DIVISION OF COMPUTER SCIENCE

Towards Usability Guidelines for Multimedia Systems

Mike Bearne
Sara Jones
John Sapsford-Francis

Technical Report No.186

March 1994

Technical Report No. 186 March 1994

Towards Usability Guidelines for Multimedia Systems

Mike Bearne, Sara Jones and John Sapsford-Francis
comrmb, comrsj, comqjs @herts.ac.uk

School of Information Sciences, University of Hertfordshire,
College Lane, Hatfield, Herts, AL10 9AB, UK

Tel: (0707) 284766, 284370, 284354
Fax: (0707) 284303

Abstract

The advent of technology which supports the concurrent presentation of information through a range of different media has raised new issues relating to the design of usable systems. While previous work in the areas of both HCI and hypermedia system usability can contribute a considerable amount to the development of such guidelines, we believe that the use of multiple output media demands an understanding of particular characteristics and limitations of users' attentional capabilities. This paper presents some initial guidelines for the design of usable multimedia systems. These guidelines are based on empirical findings regarding the nature of human attention derived from the field of experimental psychology. We believe that the provision of such guidelines for multimedia interface design will support designers in achieving the dual goals of maximising a user's flexibility in controlling the presentation of multiple concurrent media, while keeping cognitive load to an acceptable level.

Submitted for presentation at ACM Multimedia 94, the second international conference on multimedia, to be held in San Francisco, California, 15 - 20 October 1994.

Towards Usability Guidelines for Multimedia Systems

Mike Bearne, Sara Jones*, John Sapsford-Francis
M.Bearne, S.Jones, J.Sapsford-Francis@herts.ac.uk

School of Information Sciences, University of Hertfordshire,
College Lane, Hatfield, Herts, AL10 9AB, UK.

Tel: +44-707-284370
Fax: +44-707-284303

Abstract

The advent of technology which supports the concurrent presentation of information through a range of different media has raised new issues relating to the design of usable systems. While previous work in the areas of both HCI and hypermedia system usability can contribute a considerable amount to the development of such guidelines, we believe that the use of multiple output media demands an understanding of particular characteristics and limitations of users' attentional capabilities. This paper presents some initial guidelines for the design of usable multimedia systems. These guidelines are based on empirical findings regarding the nature of human attention derived from the field of experimental psychology. We believe that the provision of such guidelines for multimedia interface design will support designers in achieving the dual goals of maximising a user's flexibility in controlling the presentation of multiple concurrent media, while keeping cognitive load to an acceptable level.

1. Introduction

The advent of technology which supports the concurrent presentation of information through a range of different media has raised new issues relating to the design of usable systems. There is currently little support for the process of making or evaluating design decisions relating to these issues. The Human Factors Consultancy team at the University of Hertfordshire is, and has been, involved in the design and development of a number of hypermedia systems (see, for example, [7,8]). We are using these experiences as a basis for investigations into the development of guidelines for multimedia system design and aim, ultimately, to provide support for the use of such guidelines through formally-based tools and techniques.

In this paper, we focus on the implications of multimedia output options for system usability. Standard approaches to design for usability from the computer-human interface community (see, for example, [2]) do not yet directly address the unique characteristics of multimedia systems: while general usability criteria such as learnability, flexibility and robustness apply equally to mono- and multimedia systems, they have little to say regarding the specific benefits and drawbacks of concurrent media output. Because of historical associations with hypertext, work on usability from the hypermedia community has also tended to ignore issues relating to the use of multiple

---

* denotes primary contact

media and focus on questions of information structuring and navigation within complex hyper-structures (see, for example, [11] and [18]). We feel that an important key to understanding the design of multimedia systems is a knowledge of human attentional capabilities and the corresponding limitations which should be placed on multimedia output. In the following sections, we aim specifically to address the question of multimedia usability by drawing on basic psychological findings regarding the nature of human attention and discussing the relevance of these findings to the design of multimedia interfaces.

2. Definitions

The definitions of the terms 'multimedia' and 'hypermedia' used in this paper will be those set out by McKerlie and Preece [12]. Their definition of **multimedia** is taken from a paper by Mills [13] and is as follows: 'multimedia is an umbrella term for the integration of different elements, such as text, graphics, video, still photographs and sound, in a single application'. The 'hyper' in hypermedia is taken to suggest the notion of complex information structuring (as in hypertext), so that: 'When multiple-media are added to hypertext, the result is ..... **hypermedia**: a specialized form of multimedia.' We will focus mainly on the use of multimedia, but will briefly consider the special case of hypermedia in the final section.

We categorise output media into 5 classes:

- **Text;**
- **Graphics** including freehand or computer-generated images and video stills;
- **Videos** or sequences of graphics which change over time;
- **Sound** using music or special effects and
- **Speech** which may be pre-recorded or computer-generated.

We draw a basic distinction between what we call 'static' and 'dynamic' media. Text and graphics are referred to as **static** media, while video, sound and speech are **dynamic** in that the information they present changes over time. Video, sound and speech presentations therefore require constant attention from the user if all the information they convey is to be used.

3. Attention and Its Relevance to Multimedia Design

The use of any computer-based system involves a range of cognitive constructs and processes including perception, short and long-term memory, reasoning and problem-solving. We have chosen here to focus on attentional capabilities as we feel that it is in this area where the unique demands of using multimedia systems are most acutely felt.

The common view of attention is of a process which involves the use of mental effort in selectively processing information from a range of different sources, both internal (including our own thoughts and dreams) and external (including visual and auditory stimuli). Psychologists have been studying the nature of human attentional processes since the turn of the century and a considerable amount of literature on this subject is now available. The ability of multimedia systems to provide users with more information in more different forms than was possible with traditional monomedia technology obviously places new and increased demands on the attentional capacity of system users. It is therefore important that designers of multimedia systems should have some understanding of the characteristics and limitations of human attentional processes.

The nature of attention is still not completely understood and the literature presents a number of competing theories of attention, each of which appear to explain a range of the empirical findings gathered to date. The findings themselves are of various kinds and have varying degrees of relevance to the question of multimedia design. For example, much of the reported work on selective attention focuses on effects in very small areas (e.g. 1°) of the visual field, or on auditory tones of a particular frequency, whereas in multimedia systems we are often more interested in larger areas of the screen, or more complex sounds involving speech or music. Furthermore, the tasks investigated in research on attention are often quite simple, involving, for example, identifying words with certain characteristics from a list (see, for example, [16]), or naming letters in a briefly illuminated display (as in [3]), while for multimedia design, we may be more interested in higher-level tasks such as comprehension or problem-solving.

The following paragraphs describe phenomena on which there appears to be general agreement in the literature, and discuss the relevance of each to multimedia interface design. The main sources of information we have used in this review are [10], [4] and [17].

3.1 People are able to focus their attention on particular stimuli

We are, to a large extent, able to exercise voluntary control over the focus of our attention. This can be done on either a sensory or a semantic basis, though sensory selection of attentional focus is generally more accurate and less effortful than semantic selection.

Visual attention is often described as being analogous to a spotlight in the sense that everything within a relatively small area can be seen clearly whereas things outside the beam are hard or impossible to see. The spatial area covered by the beam apparently varies depending on the nature of the task being performed. Whatever its size, things outside the beam appear not to interfere with processing of information relating to things inside, except where the characteristics of things inside and outside are highly related, or where the characteristics of things outside are highly relevant to the viewer for some other reason. An example of the former is the Stroop effect where coloured rectangles interfere with the reading of colour names even when the two are separated by around 12° of visual field; an example of the latter is given below.

In multimedia systems - as in other window-based interfaces - our ability to focus visual attention means that there will, in general, be no problem with permitting more than one piece of static information to be displayed at any one time, as users will be able to focus their attention on each as necessary. The presence on the screen of more than one piece of information in a static medium will usually not interfere with the user's ability to process target information. Even the strongest potential interference will have little effect if the relevant pieces of information are separated by around 4 inches (a distance which corresponds roughly to 12° of visual field for viewers at a distance of 18 inches from the screen). It should, however, be remembered that in order for the concurrent display of static information to work effectively, users must be given control over the appearance and - more importantly - the disappearance of each piece of information so that they will feel comfortable with focusing temporarily on one rather than another.

Auditory attention can also be focused as illustrated by one aspect of the 'cocktail party phenomenon' described by Cherry in 1953 [1]. Cherry discovered that our ability to follow a single conversation, even when many others are going on around us, involves the use of physical differences between the relevant auditory stimuli including the sex, voice intensity and location of the speaker. As multimedia designers, this tells us that users are likely to be quite proficient at following a particular speech passage, even when others spoken in different voices are playing. Perhaps more usefully, it suggests that users are likely to have little difficulty in attending to a particular passage of speech, sound or music while other auditory information is also being presented.

Despite our normal ability to focus our attention, it may, on some occasions, be drawn involuntarily to certain stimuli, especially when they correspond to 'active schemata' or ideas of particular relevance. A familiar example is the way in which stimuli relating to the constantly active self-schema can intrude on other activities: the sound of one's own name in a conversation on the other side of a room can attract one's attention, even when it has previously been focused on something else.

From the point of view of multimedia design, the most obvious danger to avoid is the possibility of information from a previous presentation interfering with the viewing of a new one. Viewing a presentation on a particular subject will activate a particular schema in the user. When the user switches to viewing a presentation on a different subject, activation of the previous schema will die away only gradually. Information from the previous presentation remaining on the display or, worse, still being played auditorily, may interfere with processing of new information as the schema to which the old information corresponds will still be active.

3.2 People are able to attend to more than one stimulus at any one time

We know from experience that it is often possible to do more than one job or activity at the same time. For example, we may be able to drive a car while holding a conversation, or to listen to music while reading a book. Indeed, multimedia systems draw their main strength from this capability: by taking advantage of a user's ability to

read text to an accompaniment of music or other sound effects, or to view pictures while listening to a narrative, multimedia systems can convey richer information more efficiently than is possible using a monomedia interface.

Psychological research suggests that it is easier to combine different activities if:

- the activities are dissimilar, for example relying on the use of different sense modalities (e.g. Wickens et al [19] have shown that the division of attention between eye and ear is better than that for two attentional channels within either of these two modes);
- the activities are highly practised and
- the activities are simple, and do not require the use of a lot of cognitive resources.

It is not possible to translate these findings into precise recommendations for multimedia design, as the dimensions involved (similarity of tasks, degree of practice and extent of cognitive resource use) are not yet sufficiently well-defined. However, it is possible to use the above criteria as heuristics to help us decide whether users are likely to be able to attend to particular combinations of media.

An obvious combination of media which can be used effectively is that of animated graphics or video with sound, music or speech concerning the same subject. In this case, the tasks of extracting information from the graphical images and from the sound track are dissimilar (in that they rely on different sense modalities), highly practised and normally simple. Users can extract meaning from two pieces of information presented using the same modality - for example following a story described in a passage of speech while gaining an impression of mood from accompanying music - as long as the total amount of cognitive resource required is not too great (see section 3.3).

Some combinations of media are not so easy for users to tackle. Concurrent presentation of dynamic material relating to different subjects or topic areas (for example, in hypermedia systems where more than one node of information can be viewed at a time) is likely to be difficult. Comprehension of concurrently presented speech and video clips about different subjects is not feasible, as the overall level of cognitive resources required to process information about two different subjects is too great for the majority of users. Speech and video clips about the same subject may also impose too great a demand on the users processing capabilities under some circumstances: it is then the processing of information presented as speech which is likely to suffer according to Triesman and Davis [15]. Even where information presented through different modalities is about the same subject, the result may be unusable, as in a case cited by McKerlie and Preece [12] where speech and text about the same subject use different words. Speech which is difficult to understand for some reason (perhaps because it is about a difficult subject) may be made unusable by accompanying sound effects or music: for example, Salame and Baddeley [14] presented results which suggest that music - particularly vocal music - could interfere with the processing of text or verbal material. In the same way, graphics or video accompanying a piece of music which must be carefully listened to may be too distracting, and sound

effects played during a video which requires concentrated attention may cause destructive interference.

In summary, we may say that low-level tasks involving information in different modalities may, under some circumstances be performed together, but that difference in information modality is neither a necessary nor a sufficient condition for the possibility of concurrent task performance: users are able to do some tasks involving different information conveyed using the same sense modality, and are unable to do some tasks in which information needed is in different modalities. An important secondary consideration is the amount of cognitive processing involved. Attending to and processing two pieces of information concerning different subjects will always be hard, whatever the media. Even when the pieces of information concern the same subject, it may still be too hard for users to process some combinations - despite the use of different media.

3.3 Attending to information requires mental effort

While humans are, as described above, capable of attending to information from a range of different sources, it must be remembered that attending to and processing such information requires mental effort. It is therefore important to consider limiting the amount of information which users are required to process concurrently.

There are limits on attention beyond which no users should be required to operate. For example, active processing of two concurrent sources of speech output would be highly stressful for almost all users. In these cases, we may specify overall constraints to which a system as a whole must conform: for example, we may specify that a system should not at any time require its users to listen to two passages of speech at the same time, or to watch two videos at once, or to watch a video on one subject while listening to a passage of speech on another as described in the previous section.

The level of effort which is acceptable will, however, vary with individual differences between users. What for one user provides extra interest (for example, music or sound effects accompanying a video, or animations accompanying a passage of speech), may for another user seem too demanding. In cases such as these, designers must provide individual users with the flexibility to 'switch off' certain forms of output where they are not essential to the task in hand. Presentations should, of course, always be designed so that all users will be able to attend to all the information which is essential to any particular task.

3.4 People are able to direct attention to groups of stimuli sharing sensory characteristics

As suggested above, sensory selection and grouping of information is generally more accurate and less effortful than grouping on a semantic basis. Designers of monomedia systems have long been familiar with the idea that information which they intend users to link conceptually should, therefore, be perceptually linked in some way.

The extension of this idea to the context of multimedia is not straightforward. It is not clear what kind of perceptual links we could provide between information presented through different sense modalities (Would passages of speech read in a male voice be seen as more strongly linked with videos or graphics shown in darkly edged windows? Or female speech be linked with material in light-coloured windows?). However, we can go some way towards perceptually linking media directed at the same sense modality by, for example, showing text, graphics and videos relating to the same topic in windows sharing common perceptual characteristics, or having speech passages relating to the same topic spoken by the same voice.

4. Discussion

There is currently little in the way of tool-based support for designers who must make decisions about the usability of multimedia interfaces. While decisions relating to individual multimedia presentations may be made largely on the basis of intuition and common sense, the problem becomes more difficult when presentations are combined: it becomes more difficult to tell whether presentations which are usable in isolation remain usable when combined. The problem is an order of magnitude worse in the case of hypermedia systems where large numbers of nodes, each containing complete multimedia presentations can be viewed by the user in various combinations and orders.

The result of this lack of design support is that designers are often forced to make overly restrictive design decisions, simply in order to cope with the complexity of issues which would otherwise arise. For example, in hypermedia systems built at the University of Hertfordshire, it has often been decided that users should be able to view only one node in the hyper-structure at a time. This saves the designer from the difficulty of having to work out what problems might otherwise arise in the presentation of several nodes each employing various output media, but imposes constraints on the user which may not always be desirable. In some circumstances, we may wish to let users see more than one node at a time, for example, to compare their contents, or even to have a record on screen of where in the network they had come from to reach a particular point, and thus aid in navigation. In other systems, it has been decided that only one video should be viewable at any one time. While this may often be required from the point of view of hardware processing constraints, we may sometimes wish to allow users to view two videos of closely related material in parallel: perhaps to see similarities or differences between the sequences shown.

While we wish to give users the flexibility to access the combinations of media and nodes in a hypermedia network that match their particular needs, we clearly do not wish to empower users with the freedom to access combinations of media which will definitely overload their cognitive capabilities. We believe that the provision of guidelines for multimedia design based on principles of cognitive psychology such as those described above will support designers in making the tradeoffs involved in maximising flexibility while keeping cognitive load to a reasonable level.

We aim, in future, to investigate the extent to which more tool-based support for the use of guidelines such as those arising out of the discussions above (and summarised

below) can be provided. In simple cases involving the design of single multimedia presentations, we feel that tool support for the use of time-lines showing synchronization of media on different output channels (as used, for example, by Hardyman et al [6]) may be appropriate. But for the design of more complex systems, more powerful support is needed. As a first step towards providing support in these cases, we have been considering the extent to which usability requirements based on guidelines such as those below can be formally expressed, for example in CSP [9]. We aim next to determine the extent to which the formal expression of such requirements could permit proposed multimedia system designs to be verified using a computer-based theorem-proving tool such as FDR [5].

5. Summary

In this paper, we have discussed the relevance of psychological findings on attention to the design of multimedia systems. Points resulting from our discussions are intended to form a basis for guidelines for system designers concerned specifically with the usability of multimedia interfaces. These points can be summarised as follows:

- More than piece of information may be displayed on the screen at any one time using static media (text or graphics) as long as the user is given control over the appearance and disappearance of the information

- Users are likely to have little difficulty in attending to a particular passage of speech, sound or music while other auditory information is also being presented

- Information about substantially different subjects or topic areas should be removed from the display (and from auditory channels) once a user begins to study a presentation about a new subject

- When more than one form of information is presented concurrently using dynamic media, the combination will tend to be more usable if:

    - the different pieces of information are presented using different sense modalities, or require different forms of cognitive processing;

    - the medium of presentation is familiar to the user; and

    - the extraction of relevant information is easy

- Users should be given the flexibility to 'switch off' particular forms of output where they are not essential to the task in hand

- Groups of information which users are intended to see as related should be given common perceptual features

References

[1] **Cherry, E.C. 1953,** "Some experiments on the recognition of speech, with one and two ears", J. Acoust. Soc. Am. 25 : 975-79.

[2] **Dix, A. Finlay, J. Abowd, G. & Beale, R. 1993,** Human-Computer Interaction, Prentice Hall.

[3] **Egly, R. & Homa, D. 1984,** "Sensitization of the visual field", J. Exp. Psychol.: Human Percep. and Perform. 10 : 778-93.

[4] **Eysenk, M.W. & Keane, M.T. 1990,** Cognitive Psychology, LEA.

[5] **Formal Systems (Europe) Ltd. 1992,** Failures Divergent Refinement: User Manual and Tutorial.

[6] **Hardyman, L. Bulterman, D.C.A. & Van Rossum, G. 1993,** "The Amsterdam hypermedia model: Extending hypertext to support real multimedia", Hypermedia 5(1).

[7] **Hewitt, J. Sapsford-Francis, J. Bolstad, G. Eftedal, O. Halford, P. Vervenne, D. & Verheyen, M. 1993,** "MODEMA - A knowledge based browsing system to facilitate the employment of people with disabilities" in Rehabilitation Technology, Ballabio, E. Placencia-Porrero, I. & Puig de la Bellacasa, R. (Eds) IOS Press.

[8] **Hewitt, J. Sapsford-Francis, J. & Halford, P. 1993,** "Use of multimedia in a public information system" in Proceedings of the 1993 International Symposium on Multimedia Technologies and Future Applications, BCS/IEEE/RTS, Southampton.

[9] **Hoare, C.A.R. 1985,** Communicating Sequential Processes, Prentice Hall.

[10] **Johnston, W.A. & Dark, V.J. 1986,** "Selective Attention", Annual Review of Psychology, 37 : 43-45.

[11] **Kuo, F. 1993,** "A cognitive engineering-based approach to designing hypermedia applications", Information and Management 25 : 253-263.

[12] **McKerlie, D. & Preece, J. 1993,** "The hype and the media: Issues concerned with designing hypermedia", Journal of Microcomputer Applications, 16 : 33-47.

[13] **Mills, J. 1991,** Computing, 20.

[14] **Salame, P. & Baddely, A.D. 1989,** "Effects of background music on phonological short-term memory", Quarterly Journal of Experimental Psychology, 41A : 107-122.

[15] **Treisman, A. & Davies, A. 1973,** "Divided Attention to eye and ear", in Attention and Performance IV. Kornblum, S. (Ed.) New York: Academic Press.

[16] **Triesman, A. & Riley, J. 1969,** "Is selective attention selective perception or selective response? A further test", J. Exp. Psychol. 79 : 27-34.

[17] **Underwood, G. 1976,** Attention and Memory, Pergamon Press.

[18] **Waterworth, J.A. 1992,** Multimedia Interaction with Computers, Human Factors Issues, Ellis Horwood, chapter 4.

[19] **Wickens, C.D. , Sandry, D. & Vidulich, M. 1983,** "Compatibility and resource competition between modalities of input, central processing, and output: Testing a model of complex task performance", Human Factors, 25 : 227-248.