

A Multimedia Interaction Space

Frank Vetere, Steve Howard and Ying Leung

Swinburne University of Technology
 Swinburne Computer Human Interaction Laboratory
 PO Box 218 Hawthorn Vic 3122 AUSTRALIA
 {fveter, showard, yleung}@swin.edu.au

ABSTRACT Multimedia is often described simply as a combination of media. The list of possible media usually includes text, graphics, images, sound, animation, video and others. What constitutes media however is not well understood, and so the list becomes somewhat arbitrary. There is a need for a more comprehensive model of multimedia. This paper suggests 'a Multimedia Interaction Space': a three dimensional space based on the notion of media as language with a particular grammar. It distinguishes between 'base media' associated with particular sensory channels and 'composite media'. Composite media are composed of more than one base medium. The Multimedia Interaction Space provides a framework to represent multimedia interactions from an HCI perspective.

KEYWORDS multimedia interaction space, multimedia, media, mode, metaphor, language, engagement

"Our languages are our media. Our media are our metaphors. Our metaphors create the content of our culture" (Postman 1985).

1. INTRODUCTION

Given the plethora of research and development in multimedia over the last two decades (e.g. Kjeldahl 1992, Blattner & Dannenberg 1992), one would expect a well formulated body of knowledge about multimedia interactions which establishes basic principles and provides a basis for future research. Instead much of multimedia research and discussion is based on descriptions which are characterised by using more than one medium through some computer system.

A description which typifies the general understanding of multimedia is "A multimedia computer system is one that transacts or interacts in more than one medium, ie. carriers of information such as text, audio, still graphics, animation, and video" (Baecker, Grudin, Buxton & Greenberg 1995 p.400). This description of multimedia is based on exemplification rather than a robust definition.

Baecker's description is not intended to provide the foundations of multimedia research. It is a general description which reflects current understanding of multimedia. Nonetheless it captures the flavour of many multimedia applications. A computer system that uses a combination of text, audio, still graphics, animation, and video is unquestionably considered multimedia. Such descriptions have become a de facto definition.

Yet this type of description is not based on any rigorous understanding of multimedia. There is a need for a suitable structure upon which to base HCI research into multimedia.

The type of definition which views multimedia as simply an integration of text, sound and pictures is problematic. The problems arise from adopting a technology centric view rather than a user centric view to multimedia (Alty 1991, Marmolin 1992), and from conceptualising media as information carriers.

2. MEDIA AS TECHNOLOGY

When multimedia is seen as a combination of discrete media, undue emphasis is placed on the enabling technology rather than the interaction. Definitions based on audio and video representations lead to discussions about technological capabilities, such as compression techniques and transmission bandwidth. These are undoubtedly important, but from an HCI perspective they are at best only part of the picture.

The discrete media in arbitrary media lists are predominantly associated with output interactions. Animation, for example, is used only as an output medium. It has not been used as an input medium. This bias of multimedia representations towards output is in contrast to the HCI bias, whose research is largely about input. Many HCI models are concerned with input interactions (e.g. Payne & Green 1989, and Card, Moran & Newell 1983) and an analysis of interaction devices (Baecker & Buxton 1987, Helander 1991) shows a predominance of input over output devices.

Computer systems which focus on output rather than input tend to be more concerned with presentations rather than interactions (Mayes 1992). They then begin to resemble other systems that have little or no input mechanisms (e.g. television)

A technological view of multimedia does not capture the richness of the interaction. Multimedia is more than the sum of its media parts. New semantic and syntactical elements are created when media are combined (Alty 1991). When text and graphics are placed on a screen, a new representation is created. The text and graphics do not remain discrete entities.

Together they form complex conceptual links that permit new forms of interactions. We recognise this new type of interaction in print by calling it 'newspaper' or 'magazine' or 'book'. In computer systems the new representation is called 'text and graphics' or 'multimedia' – a title which makes it indistinguishable from other media combinations. If multimedia is simply a combination of media then it cannot adequately represent the structures associated with media combinations.

So when multimedia is described as being a combination of media from some media list (e.g. text, audio, still graphics, animation, and video) it becomes more concerned with the capabilities of the hardware than with the interaction from the user's point of view, it becomes more concerned with the output rather than input media and it is unable to speak accurately about the interactions resulting from the media combinations.

3. MEDIA AS INFORMATION CARRIERS

A medium is often considered the carrier of information (Blattner & Dannenberg 1992, Baecker et al. 1995). This conceptualisation of media is potentially the source of confusion and ambiguity. A carrier of information could be the text, the CD-ROM, the CPU, the VDU, the computer network or the space between the eye and the screen. If they are all information carriers then they are all media. A system which uses a floppy disk and a VDU could then be considered multimedia. This ambiguity makes it difficult to compare research results and interpret progress in multimedia. Blattner & Dannenberg (1992) acknowledge this ambiguity by accepting that a medium may or may not be of a physical form. Steinmetz & Nahrstedt (1995) get around this problem by describing six categories of media; perception, representational, presentation, storage, transmission and information exchange. Both authors attempt to manage the difficulty, but neither address the problem through a basic reconceptualisation of media.

Conceptualising media as information carrier does not provide adequate granularity. Video and sound, for

example, are both information carriers. Yet video is clearly a complex combination of many representations, including sound. Video is a 'higher order' medium. Simply considering video as another information carrier does not distinguish it from 'lower order' media. As a consequence video is often included in lists of possible media combinations (Blattner & Dannenberg 1992, Baecker et al. 1995). If all media were of equal status then adding sound to video would create a multimedia system. Again, this is misleading and ambiguous.

So media, as information carriers, leads to confusion about the nature of the information, confusion about the nature of the carrier and allows trivial media combination by not clearly discriminating between media types. A new concept of media is required.

4. MEDIA AS LANGUAGE

Alty (1996) suggests that the medium is the *language* through which the information is exchanged rather than the *carrier* of the information. He says that the valid language is defined by the restrictions of three communication mechanisms; the creation mechanism (e.g. a loudspeaker), the transmission mechanism (e.g. air) and the reception mechanism (e.g. the ear).

Conceptualising media as language is important. If media is considered language, then a grammar can be used to describe media. Discussions about the 'information' or 'information carrier' are no longer central to understanding multimedia. A simple device (e.g. a CD-ROM) cannot be considered a medium unless it is associated with a language and a corresponding grammar. 'Media as language' gives every medium a range of valid tokens (lexicon) that can be arranged in a particular ways (syntax) and that have particular meaning (semantics) depending on the context in which they are presented (pragmatics).

Alty (1991) also suggests that there be 'base' media (text, graphics, speech, sound, gesture and motion) and 'composite' media (e.g. video and animation). These two groups of media introduce a hierarchy that was lacking in Baecker's description. Combinations of base media are all valid (unlike

combinations of media from an arbitrary media-list). Meaningless media combinations, like sound and video, are not possible because video is a composite media of which sound is already one its parts.

Composite media have a particular language with emergent features that are captured by the grammar. Animation, for example, is not simply a combination of graphics and sound, but a particular medium in its own right that is fully described by its grammar.

'Media as language' deals with many of the difficulties that arise from conceptualising media as information carriers, but there is still a need for a framework and notation in which multimedia can be suitably represented.

5. MEDIA AS METAPHOR

Hutchins, Hollan and Norman (1986) developed a 'space of interfaces' in order to better understand 'directness' in direct manipulation. The two dimensions of this space are distance and engagement. Distance is concerned with the gap between the system goals and the user's goals. Engagement is based on two interaction metaphors, conversation and model world. In the interface space, direct manipulation occurs when there is a small distance between user and system goals and there is a 'feeling' of engagement. This engagement occurs when the user has a sense of direct involvement with a world of objects (model world metaphor) rather than of communicating with an intermediary (conversation metaphor).

Engagement is related to the concept of interactivity. Laurel (1992) outlines four variables which help to characterise the level of interactivity; frequency (how often one interacts), range (the number of available choices for interaction), significance (whether those choices really affect matters) and the ability to act within a representation. This last variable is the degree to which one feels like they are actually participating in the ongoing action. This is comparable to engagement. Therefore, engagement is an important measure of interactivity.

Frohlich (1992) proposes that the two interaction metaphors (conversation and model world) be used as the basis for a 'design space of interfaces'. The

conversation metaphor becomes Frohlich's language mode and the model world metaphor becomes the action mode. The 'mode' (or metaphor) is the first of four characteristics which define Frohlich's interface design space. The other three are 'channel', 'medium' and 'style'.

Frohlich defines two groups of channels; the human interface channels (voice & movement for output; hearing, sight & touch for input) and the computer interface channels (audio, visual & haptic). The term 'haptic' is used broadly to include both tactile information (specifying surface contact) and kinaesthetic information (specifying effector position).

He nominates six media; three associated with the language mode (speech, text, and gesture) and three with the action mode (sound, graphics, and motion). These are the base media that Alty (1991) used as a foundation for the composite media.

Frohlich's styles are a list of interaction techniques based on the work of Baecker & Buxton (1987)

Frohlich's use of terminology is unique. He gives a particular meaning to 'mode' which is not shared by many others. 'Mode' is often referred to as being a sensory channel (Mayes 1992), an interaction style (Blattner & Dannenberg 1992) or a computer state (Marcus 1995). Some researchers also differentiate between modality (channel and style) and mode (computer state) (Nigay & Coutaz 1993).

In Frohlich's interface design space, the interface mode or modality (as distinct from the system mode or modedness) refers to the two fundamental interaction metaphors.

The interface design space provides a detailed and comprehensive structure. It integrates media with interaction metaphors and then incorporates sensory channels and styles.

6. A MULTIMEDIA FRAMEWORK

Alty (1991) uses Frohlich's interface design space (later discussed in Frohlich 1992) to create a multimedia framework. This framework captures the concepts of interaction metaphors through the language and action modes and the concept of 'media as language' through the use of the creation and

recognition mechanisms (these are equivalent to Frohlich's 'channel').

Alty (1991) recognises that this framework could be a basis for a model of multimedia interactions. There are nonetheless, some difficulties that need to be addressed before a comprehensive model is established.

Alty's multimedia framework loses the dimension of engagement. There is no longer a continuum from the conversation metaphor to the model world metaphor as is represented by the model of interface spaces in Hutchins et al. (1986). Instead the two metaphors are represented as discrete modalities.

Alty's multimedia framework does not permit adequate representation of video or animation. The concept of composite media is introduced, but there is no structure to represent it. Alty suggests that time could be added to the graphics medium to create animation, but the framework does not currently support this type of notation.

The use of language (as a conversation metaphor) to describe one of the interaction modes is a potential source of confusion given that media is also described as language (with grammatical constructs).

A suitable model of multimedia interaction would capture the detail of Alty's framework and also address the problems concerning lack of continuum for engagement, the representation of composite media and the possible confusion with the term 'language'.

7. A SPACE OF MULTIMEDIA INTERACTIONS

It is possible now to create a three dimensional space to model multimedia interactions (figure 1). This space incorporates interaction metaphors and is based on the concept of media as language.

The three sensory channels, visual, auditory and haptic, define the three axes. The haptic channel includes both tactile and kinaesthetic information (Frohlich 1992).

The six base media occupy space along the axes; text and pictures on the visual channel, speech and sound on the auditory channel, gesture and motion on the haptic channel. Any point within the space is specified by its grammar. So, the interaction

associated with any point in the space has a particular semantic and pragmatic meaning for each of its lexical tokens and a particular syntactic structure.

The 'graphics' medium is renamed 'pictures' and includes moving as well as static pictures. A 'moving picture' is literally any graphic that moves, albeit with no accompanying sound. This could include silent film or silent animation. A moving picture has a different language structure from a static picture and so occupies a different region in the space. Video and animation can be represented in this three dimensional space by a region which includes sound, speech and pictures.

Further, the 'language' mode (Frohlich 1992) is renamed 'conversation'. This avoids possible confusion and is consistent with the earlier work of Hutchins et al. (1986) on interface spaces.

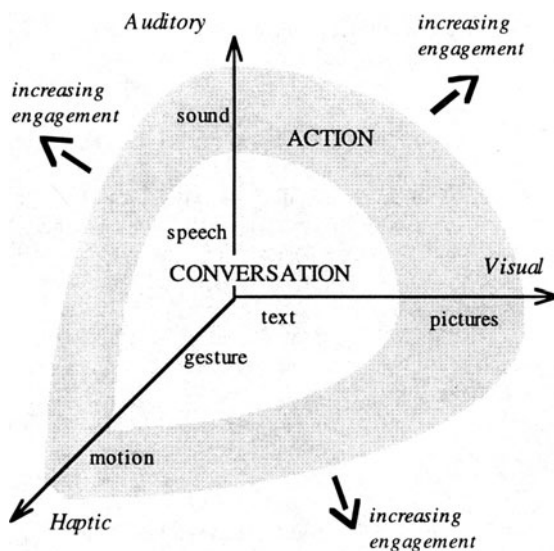


Figure 1: The Multimedia Interaction Space

The media closest to the origin (speech, text, gesture) are associated with the conversation metaphor (Hutchins et al. 1986). Those further away (pictures, sound, motion) are associated with the action or

model world metaphor. The radial distance from the origin reflects the engagement of the interaction (as defined by Hutchins et al. 1986).

The Multimedia Interaction Space (MMIS) (figure 1) permits representations of multimedia interactions and helps to clarify the concept of media. According to the model, a multimedia interaction is characterised by the language constructs (semantic, syntactic, pragmatic, lexical) of the media, by the sensory channel (the three axes) and by the interaction metaphor (radial distance from the origin)

A representation of the complete human interaction with a multimedia system requires two diagrams, one for input and the other for output. These two diagrams would represent the nature of multimedia interaction (both input and output) and the level of engagement for any particular computer system. The complete interaction would be represented by spaces in both the input and output diagrams.

At any particular time, the interaction is represented by a point or region in the space. At some other time, the interaction will occupy another place. It is theoretically possible to show how the interaction changes over time. The conglomeration of these interactions produces a diagram which is a representation of the entire interaction over time. This diagram could be considered a 'map' of the interaction.

It is significant that engagement is included in the MMIS model. Engagement is an important feature of interactivity and interactivity is often associated with multimedia. This model of multimedia interaction captures three of the Laurel's (1992) four variables; frequency (amount of interaction), range (the number of available choices for interaction) and engagement.

It is also possible to use the model to represent new types of multimedia interactions that have yet to be fully exploited or developed. This is particularly relevant for media along the haptic channel.

The model is presently restricted to three of the five human sensory channels. It is conceivable that a five dimensional space could be created if there were a computer system that made use of the gustatory or olfactory senses.

8. APPLYING THE MODEL

If extended, the MMIS model could be used to represent input or output interactions of a computer application. The Encarta® encyclopaedia by Microsoft® is a typical example of a multimedia product which could be represented using the MMIS model (figure 2).

Two diagrams are required to represent the input and output interactions, but only the output diagram is presented here. Encarta's output uses media along two channels. It does not use haptic media, so the output media interaction can be represented along two dimensions, auditory and visual.

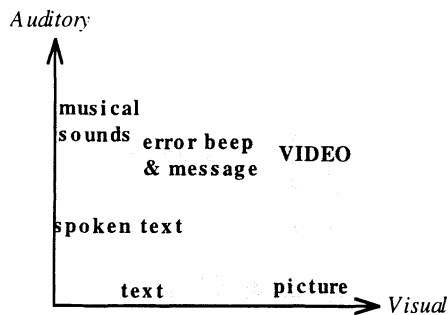


Figure 2: Using MMIS to represent the output media interaction of Microsoft® Encarta®

The MMIS model is used to represent the media associated with the output interaction. It is not a complete representation of the interaction. It is only intended to exemplify the model. Not all the media combinations have been included.

The diagram also makes many assumptions about the nature of the media and the level of engagement. A complete and thorough representation requires operationalisation of the grammar for each media and for the characteristics of engagement. This would provide a basis on which to determine the size and position of the regions in the space. Presently, the model offers a framework on which operationalised characteristics may be built.

The Multimedia Interaction Space provides an elegant mechanism for representing the use of media in a computer application. As alluded to in figure 2,

interactions with different multimedia systems would yield different MMIS maps. For example, short musical compositions would be differentiated from other sounds such as 'beeps' (which may contain textual error messages) and video (which contains moving pictures). In principal, the MMIS model supports descriptions of media types in a computer system and allows meaningful comparisons between systems.

9. CONCLUSION

The Multimedia Interaction Space makes progress towards a better understanding of multimedia interactions. The model provides some immediate benefits. These include:

- replacing general definitions of multimedia with a structured model which permits comprehensive representations of multimedia interactions
- reconceptualising media as language rather than as an information carrier
- allowing input as well as output media representations
- distinguishing between base media and composite media
- representing multimedia interactions from a user perspective by incorporating interaction metaphors
- combining the concepts of engagement and interactivity within a model of multimedia interaction.

The model also provides other advantages that could be exploited with further research. These include:

- being unconstrained by existing media representations and allows for the representation of undeveloped media (especially using the haptic channel)
- the possibility of representing changing interactions over time
- providing a potentially calculable conceptualisation of multimedia interactions

More work needs to be done in operationalising the concepts and in representing the possible relationships between the shaded areas. Nonetheless, the model makes a contribution towards creating a comprehensive structure upon which to base HCI

research in multimedia. It does not explicitly represent interaction styles, as in Frohlich's (1992) design space, but it does attempt to integrate the key concepts of language and metaphor within a multimedia framework.

The integration of language and metaphor with multimedia echoes Postman's (1985) aphorism which prefaced this paper. He suggests that the use of language and metaphors through media has cultural implications. This is true for multimedia which is having a significant effect on society and culture. The Multimedia Interaction Space begins to acknowledge the role of language and metaphor in multimedia interactions.

10. REFERENCES

- Alty, J. L. (1996) "Multimedia Interface Design" Tutorial notes at OzCHI 96
- Alty, J. L. (1991) "Multimedia - What is it and how do we exploit it?" in Diaper, D. & Hammond, N.(Eds.) *People and Computers VI: Proceedings of the HCI '91 Conference* Cambridge University Press, Cambridge UK, pp 31-41
- Baecker, R. M., Grudin, J., Buxton, W. A. S. & Greenberg, S. (1995) *Readings in Human-Computer Interaction: Towards the Year 2000 (second edition)* Morgan Kaufmann, San Francisco CA.
- Baecker, R. & Buxton, W. A. S. (1987) *Readings in Human Computer Interaction* Morgan Kaufmann, Los Altos CA
- Blattner, M. M. & Dannenberg, R. B. (1992) *Multimedia Interface Design* Addison Wesley, Reading MA
- Card, S. K., Moran, T. P. & Newell, A. (1983) *The Psychology of Human-Computer Interaction* Lawrence Erlbaum, Hillsdale NJ
- Frohlich, D. M. (1992) "The Design Space of Interfaces" in Kjeldahl, L. (Ed.) *Multimedia: Systems, Interaction and Applications*. Springer-Verlag, Berlin pp 53-69
- Helander, M. (1991) *Handbook of Human-Computer Interaction*. Elsevier Science, Netherlands
- Hutchins, E. L., Hollan, J. D. & Norman, D. A. (1986) "Direct Manipulation Interfaces" in Norman, D. A. & Draper, S. W. (Eds.) *User Centered System Design*. Lawrence Erlbaum, Hillsdale New Jersey pp 87-124
- Kjeldahl, L. (1992) *Multimedia: Systems, Interaction and Applications*. Springer-Verlag, Berlin
- Laurel, B. (1992) *Computers as Theatre*. Addison Wesley, Reading MA
- Marcus, A. (1995) "A Comparison of Graphical User Interfaces" in Baecker, R. M., Grudin, J., Buxton, W. A. S. & Greenberg, S. *Readings in Human-Computer Interaction: Towards the Year 2000 (second edition)* Morgan Kaufmann, San Francisco CA, pp 457-468
- Mayes, J. T. (1992) "The 'M-Word': Multimedia Interfaces and Their role in Interactive Learning Systems" in Edwards, A. D. N. & Holland, S. (Eds.) *Multimedia Interface Design in Education*. Springer-Verlag, Berlin pp 1-22
- Marmolin, H. (1992) "Multimedia from the Perspective of Psychology" in Kjeldahl, L. (Ed.) *Multimedia: Systems, Interaction and Applications*. Springer-Verlag, Berlin pp 39-51
- Nigay, L. & Coutaz, J. (1993) "A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion" in *Proc. INTERCHI '93 Human Factors in Computing Systems* ACM Press pp 172-178
- Payne, S. & Green, T. R. G. (1989) "Task-action grammar: the model and its developments" in Diaper, D. (Ed.) *Task Analysis for Human-Computer Interaction* Ellis Horwood, Chichester
- Postman, N. (1985) *Amusing Ourselves to Death*, Heinemann, UK
- Steinmetz, R. & Nahrstedt, K. (1995) *Multimedia: Computing, Communications and Applications* Prentice Hall, Upper Saddle River NJ