The media in multimedia for learning: differentiating form from content, technology and domain

Wendy Doubé
Multimedia Department
Swinburne University
Australia
wdoube@swin.edu.au

Abstract: Whereas the polysemous term media is unambiguous within a single domain, multiple ambiguous interpretations can now be found in discussions of electronic convergent mass and communication media. The different interpretations of media in multimedia can be broadly classified as convergent when considering the end product, such as a news web site, or divergent when considering discrete elements, such as text, audio and video. The latter interpretation is commonly used in higher education multimedia applications. This paper revisits theories of multimedia in learning with reference to current standards for learning object metadata to produce a simple model of divergent or discrete media in multimedia. The model clearly distinguishes media form from content, technology and domain with the aim of reducing ambiguity.

Introduction

The term medium is defined broadly as ‘something in a middle position’ or ‘a means of effecting or conveying something’ (Dictionary, 2010). Consequently the usage of its plural, media, is polysemous to some extent, differing across domains. For example, in art it implies the materials used in the creation of an artwork. In journalism it has traditionally described mass communication broadcast media such as newspapers, television and radio. Likewise, definitions of the noun and adjective multimedia are somewhat fluid depending on a focus that ranges from sensory channels of communication such as sight, hearing and touch to physical delivery such as devices and file formats. From the perspective of computing topics it can be defined as “a combination of text, data, pictures, sound, video, etc., as on a CD-ROM compact disc, for interactive access through electronic computers” (Webster's New World College Dictionary, 2010) or “the integration of multiple forms of media. This includes text, graphics, audio, video, etc “ (The Tech Terms Computer Dictionary, 2010). Other media mentioned in the definitions of multimedia include print, animation (Webopedia, 2010), hypertext (Oxford Dictionaries Online, 2010), speech and music (Hoogeveen, 1997). Similarly, in learning and cognition multimedia research, the media in multimedia are generally categorized as representation formats such as text, images, and sound. Heller and Martin (1999) developed a multimedia taxonomy with three dimensions: media type, media expression and context. Media type lists in order of increasing complexity: text, sound, graphics, video and multimedia, where multimedia is the ‘seamless’ combination of two or more of the other media.

The generic nature of this interpretation has two main benefits. Firstly, it can be applied across domains regardless of changing delivery devices and formats. Secondly, media can be easily mapped to user visual and auditory sensory channels and models of cognition. In other words, media can be distinguished from content and technology and it can be mapped to form. Consequently this interpretation has been adapted by some metadata standards for learning objects, with metadata approaches replacing taxonomies to some extent, possibly in recognition of the complexity of multimedia in learning environments. In the W3C Synchronized Multimedia Integration Language (SMIL 3.0) Recommendation Report (SMIL 3.0, 2008), generic media object elements have the synonyms of animation, audio, image, text and textstream (dynamic text) In the Sharable Content Reference Model (SCORM 2004 4th Edition), content model assets (learning resource building blocks) are defined as electronic representations of media such as text, images or sound (Advanced Distributed Learning (ADL), 2009). Alternatively, the Dublin Core Metadata Initiative Metadata (2010) Metadata Element Set Type Vocabulary distinguishes Format (physical attributes such as file type) from Type (nature or genre). However Type does not distinguish media (image, moving image, sound, still image, text) from other resource types.
Convergent use of the terms media and multimedia

Electronic technology has changed the nature of creative, mass and personal communication. Consequently, technology is being incorporated into the content of higher education courses that traditionally covered Art, Journalism and Communication and are now possibly being rebadged with names such as New media, Digital media, Multimedia, Integrated media, Interactive media, Transmedia (Huang, 2009). Whereas the meaning of the term media may have been unambiguous within a single domain, it is now intertwined and possibly confused with meanings from different domains. Social media, rich media, new media are referred to in literature that simultaneously refers to the Internet, magazines and film as media. The ambiguity or polysemy between different domains is apparent when multimedia is considered a feature of online media, together with interactivity and hypertext (Deuze, 2003, 2004). The dichotomy is explicitly stated when multimedia is defined as “the combined action of multiple media like digital newspapers, e-mail, blogs and fora instead of the combination of different sign systems like text, photo and video that are in fact nothing more than the materials the different online media are built of” (Opgenhaffen, 2008). In fact the first definition is an attempt to adapt traditional definitions of mass media to online technology and the second is a fairly accurate definition of media as discussed in computing domains, in keeping with the broad definition of a medium. The first definition approaches the final combination of media holistically and does not distinguish form from content, technology, or domain. In contrast, the second definition focuses on the attribute of media form and filters out technology and content from the mass media in newspapers, letters, diaries (static text and images) and in group discussions (sounds) and considers them as translated for the technology of online delivery (static, dynamic, hyperlinked text and images). The similarity between the polysemous interpretations of the terms media and multimedia could promote assumption and misinterpretation.

Together with semiotics and technology, content clearly plays a major role in the different interpretations. In mass communication literature, discussion of content can be indistinguishable from that of delivery format or representation. This trend appears to have crossed the boundaries into domains that traditionally separate content from form. In 2005, Wikipedia defined multimedia as “the use of several different media to convey information (text, audio, graphics, animation, video and interactivity). Multimedia also refers to computer media” (Wikipedia, 2005). In 2010, Wikipedia includes content in its definition of multimedia as “media and content that uses a combination of different content forms. Multimedia includes a combination of text, audio, still images, animation, video, and interactivity content forms” (Wikipedia, 2010). This definition of multimedia leaves media undefined. Furthermore, the hyperlink from ‘content form’ leads to a page titled ‘Content format’ (not form), primarily concerned with data encoding and mapping, at the delivery end of the multimedia definition range and consequently not easily mapped to sensory channels or generalized to learning objects. In other words, the current Wikipedia definition of multimedia does not separate media form from content or technology.

With increased broadband capacity, multimedia is being incorporated into educational environments at an unprecedented rate. Students are accessing online materials including audio files and videos as a matter of course. Many are participating in educational immersive environments and using social networks and wireless devices such as mobile phones and iPads as tools for learning. At this stage of convergence of the different domains that discuss multimedia and given this conference context of educational media, it is useful to revisit and refine the interpretation of the term media as it applies to multimedia learning.

Perspectives of media in multimedia learning

In early investigations of film and television in learning Salomon (1979, p. 26) did not specifically define educational media but referred to it as complex entities consisting of technology, contents, instructional situations and symbol systems paradoxically not equated with any of these dimensions. He argued that a medium is best distinguished by the symbol systems it uses because symbol systems best represent the domain of...
learning and best match mental models. Attributes of a medium, not the medium itself, affect learning and different media attributes cater to the internal representation systems of different learners. In early investigations of multimedia learning, Reeves distinguished media "all means of communication, whatever its format" from technology "any object or process of human origin that can be used to convey media" (Reeves, 1998). According to this definition, text can be regarded as a medium and books can be regarded as an equivalent form of technology to computing systems. Conversely, DVD players and televisions are viewed as technological devices that convey electronic information, and not as media. Both these early approaches make valuable distinctions for the current analysis of media in multimedia.

In possibly the most comprehensive body of multimedia learning research, Cognitive Theory of Multimedia Learning (CTML), multimedia is defined simply as a combination of words and pictures (Mayer, 2005). This body of work investigates the impact of multimedia on learning from the perspective of cognitive models, in particular the limited resources of working memory. Cognitive load theory proposes a working memory that holds very few elements of information for an extremely brief period of time and can be overloaded when they compete for its limited resources (Sweller, 1994). Dual-coding theory proposes that working memory can be expanded if its visual and auditory channels are simultaneously employed when words are heard and pictures are seen (Paivio, 1986). CTML draws from both theories to assert that when words are simultaneously heard and read, working memory can be overloaded during attempts to organise the two sources of information into a verbal model.

Also working within the cognitive load framework, Tuovinen (2001) classified nine Learner-Content Interaction Media presentation dimensions (Table 1) in the Distance Education Four Interaction model. Each category can contain the categories in the sequence to its left, bearing in mind that verbal and pictorial media compete for working memory resources. For example, graphics can incorporate text as well as pictures. Video can contain graphics and text. Sound is presented separately, to emphasise that working memory can be expanded when both auditory and visual channels are employed. Each combination of visual media and sound is viewed as a separate category.

<table>
<thead>
<tr>
<th></th>
<th>Text</th>
<th>Graphics</th>
<th>Video</th>
<th>Virtual Reality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound</td>
<td>Text &amp; Sound</td>
<td>Graphics &amp; Sound</td>
<td>Video &amp; Sound</td>
<td>Virtual Reality &amp; Sound</td>
</tr>
</tbody>
</table>

Table 1: Nine learner-content interaction media dimensions (Tuovinen, 2001).
Electronic technology is not included in all definitions of multimedia (Mayer, 2005; Reeves & Nass, 1996; Schnitz, 2005). For example, print, in contrast to text, is sometimes referred to as a medium. Electronic multimedia supports media attributes not available in non-electronic media. Electronic technology can record and transmit sound as well as moving images and text, encoded as digital or analogue. A major feature of digital electronic technology is interactivity. From the perspective of journalism and communication, Deuze (2003) considers three categories of multimedia interactivity: navigational by moving around a site; functional by interacting with other users or producers of a site, for example, in fora or by submitting a web-based form; and adaptive where user actions alter the content of the site, for example by wiki’s, content management systems or adaptive software.

From the perspective of computing, although some of the media definitions and models discussed above include both static and dynamic, or moving, images (animation, video) they do not address interactivity. As recognized by Deuze (2003), at the most basic level of interactivity multimedia facilitates navigation from one view or screen to another. Hyperlinks can be considered to be an important feature of digital multimedia, enabling navigation interactivity. This type of multimedia interaction is recognized in Level 6 of Tuovinen’s (2001) eight two-way multimedia learner-content interactions (Table 2), some of which can be found in contexts other than learning, such as filling out a web-based form. The other interactions consist of multiple media and media attributes.

<table>
<thead>
<tr>
<th>Level</th>
<th>Learner-multimedia content</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Passive, 2-way</td>
</tr>
<tr>
<td>2</td>
<td>Choices from a hierarchy</td>
</tr>
<tr>
<td>3</td>
<td>Information update control</td>
</tr>
<tr>
<td>4</td>
<td>Construction with components</td>
</tr>
<tr>
<td>5</td>
<td>Participation in simulation</td>
</tr>
<tr>
<td>6</td>
<td>Navigation of hyperlinked information</td>
</tr>
<tr>
<td>7</td>
<td>Operation in a microworld</td>
</tr>
<tr>
<td>8</td>
<td>Multimedia creation</td>
</tr>
</tbody>
</table>

Table 2: Eight levels of 2-way interaction with multimedia (Tuovinen, 2001)

Also considering multimedia interactivity from a pedagogical perspective, Laurillard (2002, Part II) proposed a conversational framework for learning based on the premise that learning proceeds from dialogue between the instructor and learner in which the learner’s perception of the content gradually converges with that of the instructor. She asserted that learning from multimedia is predicated upon the degree to which it could support the conversational framework. She identified five forms of educational media delivered as multimedia, according to their ability to conform to the steps in the conversational framework. Some of these media forms could be classified as individual media (interactive hypermedia, narrative print), some as devices and some as software applications.

From the perspective of learning object metadata standards, the Draft Standard for Learning Object Metadata defines Educational Category Section 5.1 Interactivity as three interactions – active (two way), expository (one way), and mixed. Section 5.2 Learning Resource Types are listed as exercise, simulation, questionnaire, diagram, graph, index, slide, table, narrative text, exam, experiment, problem statement, self assessment and lecture (IEEE, 2002). The elements in both sections are independent of media. However, the description of interactivity in Section 5.1 lists text, images and sound, as well as essays, video clips, all kinds of graphical objects and hypertext documents. In order to separate form from content, this paper would prefer to distinguish between narrative in text, images and sound in Section 5.1 and would replace the term ‘essays’ with text documents in Section 5.2.
A general model of media in multimedia learning

The purpose of this section is to develop a simple taxonomy of the media in multimedia, that like Salomon’s (1979) symbol systems, can be generalized to learning objects in any technology or format and can also be mapped to cognitive structures and sensory channels. Each medium and combination of media should be capable of being separated from the content that it conveys and the technology of its delivery. The taxonomy will be based on Tuovinen’s (2001) taxonomy of individual media and will take into consideration:

a) the CTML definition of multimedia as words and pictures (Mayer, 2005)
b) the auditory and visual channels in the model of working memory (Paivio, 1986)
c) Salomon’s assertions that media attributes, not the medium itself, effect learning (Salomon, 1979)
d) the terminology in the learning object standards discussed above (Advanced Distributed Learning (ADL), 2009; Dublin Core Metadata Initiative Limited, 2010; IEEE, 2002).

Visual media in the model

The category of Virtual Reality in Tuovinen’s (2001) taxonomy of individual media (Table 1) is a broad and complex concept and will not be included in the model. The remaining media can then be categorized by visual and auditory channels. Working firstly with the visual channel we can deduce that Graphics are a type of Image and Video is another type of Image, with the attribute of dynamic motion. Likewise Text can be static, as in print, or dynamic, as in PowerPoint presentations. We can therefore re-categorise the visual media with the attributes of static and dynamic:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Visual Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static</td>
<td>Still Text</td>
</tr>
<tr>
<td></td>
<td>Still Image</td>
</tr>
<tr>
<td>Dynamic</td>
<td>Moving text</td>
</tr>
<tr>
<td></td>
<td>Moving Image</td>
</tr>
</tbody>
</table>

Table 3: Static and dynamic attributes of visual media

This approach is sufficiently general to conform with CTML while encompassing a broader range of images than graphics and video. For example, still images could be further categorized as photographs, graphics and diagrams, all of which can contain text. They could have other attributes of analogue or digital electronic or they could be non-electronic. Moving images could be further categorized as electronic analogue or digital films, animations or videos. The ‘contains’ relationship would operate diagonally. For example, moving images can contain moving text, still images and text.

Pedagogical multimedia interactions can incorporate multiple elements in visual media. Consider Tuovinen’s (2001) Level 2 Choices from a hierarchy interaction implemented in an online quiz (Table 2). Images could be present in the form of checkboxes that could be dynamic when changing from unchecked to checked. Text could be present in instructions, in the question, in the feedback supplied to students, and possibly in text boxes into which students enter an answer. It could appear in labels, headings, and in the buttons that students click. Regardless of the interaction, it is still text, whether static or dynamic. This learning interaction consists of several stages and components, some of which could be considered dependent on technology. Furthermore some components are dependent on content and could be independent of medium. In contrast, the interactivity of Tuovinen’s (2001) Level 6 Navigation of hyperlinked information interaction (Table 2) hypertext (text with a hyperlink to another location) and hypermedia (text and images hyperlinked to another location) could be seen as an attribute of a text or image element, independent of content or technology other than hyperlinks. The SCORM Asset definition refers to both media clips and lines of Hypertext Markup Language (HTML) as basic learning resource elements, in contrast to complex interactive resources (Advanced Distributed Learning (ADL), 2009). The draft IEEE standard includes HTML in the list of media in the description of interactivity (IEEE, 2002). Even interactivity as viewed from the perspective of online journalism recognizes a navigation dimension of interactivity, albeit distinct from hyperlinks (Deuze, 2003). Consequently the model of visual media will also include an additional attribute of hyperlink or navigational interactivity (Table 4).

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Visual Medium</th>
</tr>
</thead>
</table>

- 2011 -
Table 4: Static, dynamic and interactive attributes of visual media

**Auditory media in the model**

The CTML model of working memory distinguishes words that are heard from words that are read – verbal sounds from visual text (Figure 1). Non-verbal sounds are also used in learning environments. Non-verbal sounds like instrumental music can also incorporate words. Although non-verbal sounds are not always distinguished from verbal sounds in models of multimedia learning, or in the physical file types in the different learning object metadata standards, they convey very different information and have very different pedagogical implications. Consequently our model of auditory media will distinguish between verbal and non-verbal sounds. Figure 2 shows the final model of visual and auditory media in multimedia. Each combination of the six categories of visual media can be combined with both verbal and non-verbal sounds.

![Figure 2: The model of visual and auditory media in multimedia](image)

**Applying the model**

If we apply the model to a common example of a web page for a subject in a higher education learning management application such as Blackboard, the main content of the page would be in static text. It might have a static hypermedia image of the lecturer linked to an audio clip containing the verbal sounds of the lecturer welcoming students to the class. This example illustrates how the model differentiates the media used to convey content from the technology of an audio clip embedded in a web page, and from the content of the page and the message delivered by the lecturer. The same content could be adapted to other media and technologies. At the same time, the model is congruent with sensory and mental models of cognition and learning. From the perspective of CTML, the verbal sounds of words spoken by the lecturer would utilize the auditory channel to be organized into a verbal model while the image of the lecturer would utilize the visual channel to be organized into a pictorial model, thus expanding working memory to enable processing. The text content would utilize the visual channel and would be organized into a verbal model without competition from sounds. Furthermore, the example could be described by the resource elements of any of the metadata models discussed in the Introduction section of this paper, and integrated into their technical and instructional elements.
Expanding and adapting the model

The model could be expanded in at least three main directions. Firstly, research into tactile media in learning could illuminate a third dimension. Secondly each category of media could be further defined, for example along electronic/non-electronic, analogue/digital dimensions. Dynamic images could be categorized as film, video, animation etc. Thirdly, higher levels of interactivity beyond navigational hyperlinks could be incorporated into hierarchies that group the individual media into more complex aggregations. Alternatively, considering that the term media is widely used in its holistic or convergent interpretation as applying to mass and communication media, it could be replaced by another term such as mode or modality (MacGregor, 2003), or qualify another noun such as type (Dublin Core Metadata Initiative Limited, 2010) or representation (Advanced Distributed Learning (ADL), 2009) when used in a discrete or divergent interpretation in a multimedia computing domain (Opgenhaffen, 2008).

Conclusion

With electronic technology changing the nature of domains such as art, journalism and communication, multiple meanings of the polysemous terms media and multimedia can now be found intertwined and confused in the same document. In a single document it can be found in a holistic or convergent sense describing mass and communication media categories such as social and rich media, mass media like magazines, communication media like email or blogs and also divergent or discrete media like text and video. This paper has revisited definitions and taxonomies of multimedia in cognition and learning to establish some key features of its current usage in those domains. It has also examined corresponding elements in current models of metadata for learning objects. Based on this literature it has proposed a simple model of visual and auditory media in multimedia that separates form from content and technology while enabling congruence with current standards of metadata for learning objects. The model promises a basis for expansion into more detailed hierarchies and for adaptation to media used by other senses such as touch.

References


Webopedia. (Ed.) (2010) Free online dictionary for words, phrases and abbreviations that are related to computer and Internet technology. QuinStreet.
