Editor: Glorianna Davenport
MIT Media Lab

Defining Multimedia

Helen Purchase
The University of Queensland

Multimedia is variously and often ambiguously defined. While most people might accept "a mix of [voice, text and graphics],"[1] they might resist calling a live lecture on a titled work of art a multimedia presentation.

On the other hand, many definitions focus entirely on technology: "Multimedia seems to be defined by the hardware required … rather than by the user's experience."[2] For example, despite the statement that "any computer application that employs a video disk, images from a CD-ROM, uses high-quality sound, or uses high-quality video images on a screen may be termed a multimedia application,"[3] I doubt that anyone would use the term multimedia for a computer application that merely plays a piece of music.

In this article I suggest a model of media objects that does not refer to technology or interactivity, but rather concentrates on the nature of the text. This model provides a useful basis for defining multimedia communication securely and unambiguously.

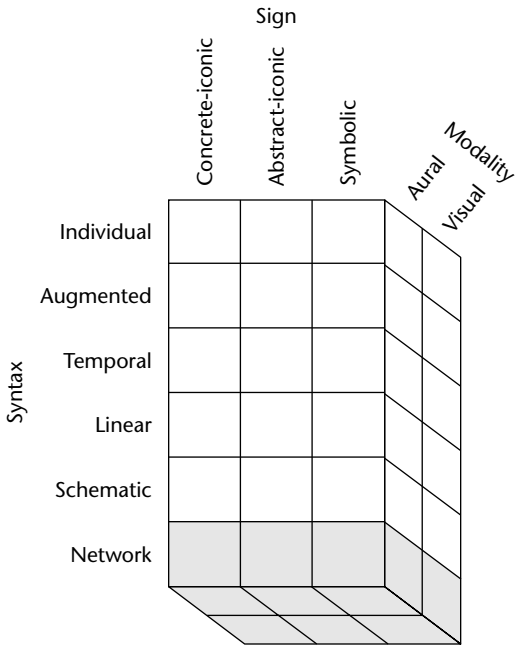

Figure 1. The model's three dimenions: sign, syntax, and modality.

Semiotic terminology

This article uses Peirce's semiotic definitions and categorizations; that is, a sign is an intimate relation between an object (term) and an interpretant (concept).[4] While the nature of the bond between a sign's two components is irrelevant, a symbol is a particular category of sign where the relationship between its object and interpretant is arbitrary. A semiotic system (code) organizes patterns of particular signs (usually rule- and convention-based) that comprise a system of meaning, and a symbolic system is a type of semiotic system based on symbols. A semiotic system thus consists of a syntax that defines the manner in which terms may be organized and a semantics that indicates how meaning can be attributed to a syntactically correct pattern of terms.

A message is defined as a syntactically correct and meaningful combination of terms in a semiotic system, and text (representation) as its physical realization. A representational system (or medium) is an abstract term that refers to the physical realization of the rules and conventions that comprise a semiotic system, and a device is a physical object used for communication via a semiotic system. Communication occurs when text is created according to a particular code and transmitted via a device. The receiver of the text decodes it to extract meaning, choosing to impose a particular code on the text to interpret it.

For example, music is a semiotic system consisting of signs (notes), where each note consists of a term (such as E flat) and a concept (the sensation produced by the corresponding sound wave). A figure from Beethoven's ninth symphony is a message, the score is a text, an orchestra is a device for communicating it, and musical notation is the abstract term denoting the corresponding representational system.

A taxonomy of representational systems

I propose a model based on three dimensions: the nature of the sign, the arrangement of the signs, and the modality (see Figure 1).

The first dimension: the nature of the sign

Bruner's classification identifies three different types of semiotic systems: enactive, iconic, and symbolic.[5] The enactive system is based on physical movement and learned responses (such as the actions required for riding a bicycle), the iconic system depends on imagery and perception (like pictures or photographs), and the symbolic system uses symbols that do not have a perceptual relationship with the concepts they signify (for example, words or a red traffic light).

I do not include the enactive category in the model proposed here, as its physical nature is more concerned with personal action than with communication.

Here I propose a further division of the iconic category, based on the object's nature. Arnheim categorized visual signs (images) as pictures or symbols, depending on their level of abstraction:[6]

> An image serves merely as a sign to the extent to which it stands for a particular content without reflecting its characteristic visually.... Images are pictures to the extent to which they portray things located at a lower level of abstractness than they are themselves. They do their work by grasping and rendering some relevant qualities—shape, color, movement—of the objects or activities they depict.... An image acts as a symbol to the extent to which it portrays things which are at a higher level of abstractness than is the symbol itself (pp. 136-138).

If icons occupy one end of an "abstraction continuum" and symbols the other, a further category of sign—proposed here—falls between the two, classifying icons as either concrete or abstract.

Icons perceived as identical to the concept that they represent (like photographs) have a very low level of abstraction, hence are here termed concrete-icons. Icons sensed as similar, but not identical, to the concept lie higher up the abstraction scale. These abstract-icons are not as abstract as symbols, since a perceptual relationship still exists between object and concept, but neither are they perceived as identical to the concept.[7] An example of an abstract-icon is a road sign warning against falling rocks in the area.

The three values for the first dimension of the model proposed here therefore relate to the nature of the sign: concrete-iconic, abstract-iconic, and symbolic.

The second dimension: the arrangement of the signs

Original semiotic classifications were based on simple communication devices like paper or static displays. The second dimension for the model takes into account the increasing use of technology and consequently the more complex manner in which texts may be communicated. This dimension adds the organization (or syntax) of the representational system.

To communicate their meaning correctly, the objects in representational systems need to be arranged in a particular manner. I propose five syntactic methods of arranging icons and symbols: individual, augmentation, temporal, linear, and schematic.

- *Individual*: The simplest arrangement, having only a single object to communicate a single concept and only a single moment in time required to receive the message.
- *Augmentation*: An extension to the individual arrangement. Like the individual syntax, only a single object is used, augmented with one or more additional features that contribute additional meaning to the sign's interpretation.
- *Temporal*: Time is important to this arrangement—the message cannot be interpreted if only a snapshot is taken. Like the individual syntax, only a single concept is communicated by a single object, and the interpretation does not change over time.
- *Linear*: Linear systems place the objects in a purely sequential manner, and interpretation of the message depends on the objects being considered in this linear arrangement over time. However, unlike the temporal arrangement, more than one concept may be communicated. This is similar to Heller and Martin's definition of "temporal,"[8] which includes the notion of the message's content changing over time.
- *Schematic*: Schematic representational systems use spatial indicators to show the information's structure. They represent relationships between the concepts associated with the individual objects in a 2D or 3D manner according to a conventional code.

Table 1. Visual modality examples.

	Concrete-Iconic	Abstract-Iconic	Symbolic
Individual	Any photograph	An iconic road sign (see Figure 4)	Any written word
Augmentation	A shaped photograph, such as a star-shaped photo of a popular singer (see Figure 2)	Road signs whose color provides additional information (see Figure 5)	A word whose font provides additional information, such as a fast-food restaurant called *Express*, where italics imply speed
Temporal	A continuous rolling film, such as a film of a waterfall	A repeating sequence of drawings, such as the continuous changing background in a cartoon used to indicate that an object is falling (see Figure 6)	A repeated symbol, such as a rotating cursor
Linear	Any film	A sequence of drawings, such as a cartoon strip	A sequence of written words such as a paragraph
Schematic	A taxonomic diagram, such as a diagram depicting the management hierarchy of an organization, showing the relationship between concepts that are represented as photographs (see Figure 3)	An iconic chart, such as a bar chart using icons of people to represent numbers (see Figure 7)	A text where the 2D or 3D spatial layout of the symbols is significant, such as a desktop interface

Figure 2. A shaped photograph illustrating the use of a concrete-iconic augmented text. The shape adds information about the status of the person in the photo, in this case a popular singer.

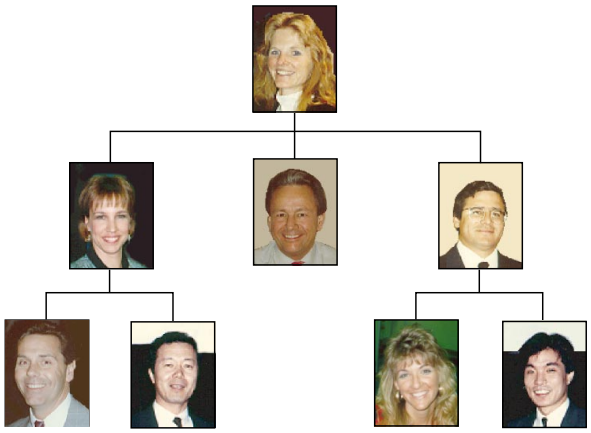

Figure 3. A management hierarchy diagram illustrating the use of a concrete - iconic schematic text. Instead of labels indicating the people in the organization, photographs are used.

The third dimension: the modality

The third dimension requires a clear distinction between the oft-confused terms multimedia and multimodal. In its common use, multimedia refers to the nature of the text used in communication, both as output from a technological system (video, sound, and graphics) and, less obviously, the human input (touch and speech). Multimodal, however, relates specifically to the senses used by the receiver of the text (visual, auditory, and tactile).[9]

This third dimension in the model therefore has two values—visual and aural—relating to the two senses most commonly used for communication.

Examples

Having defined the three dimensions to classify texts, validating these dimensions requires associating all cells in the resulting model with an existing representational system. Distinct examples of media types exist for all 30 cells. Table 1

Figure 4. A road sign illustrating the use of an abstract-iconic individual text. In this case, the sign warns of falling rocks on the road ahead.

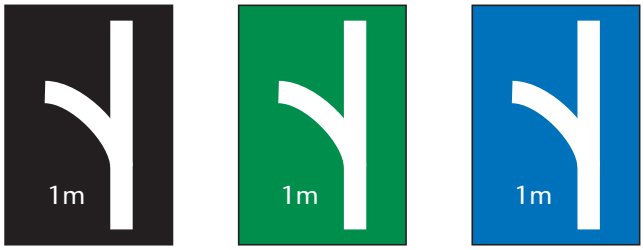

Figure 5. Three road signs, with color indicating whether the road is local, national, or a motorway. This example illustrates the use of an abstract-iconic individual text.

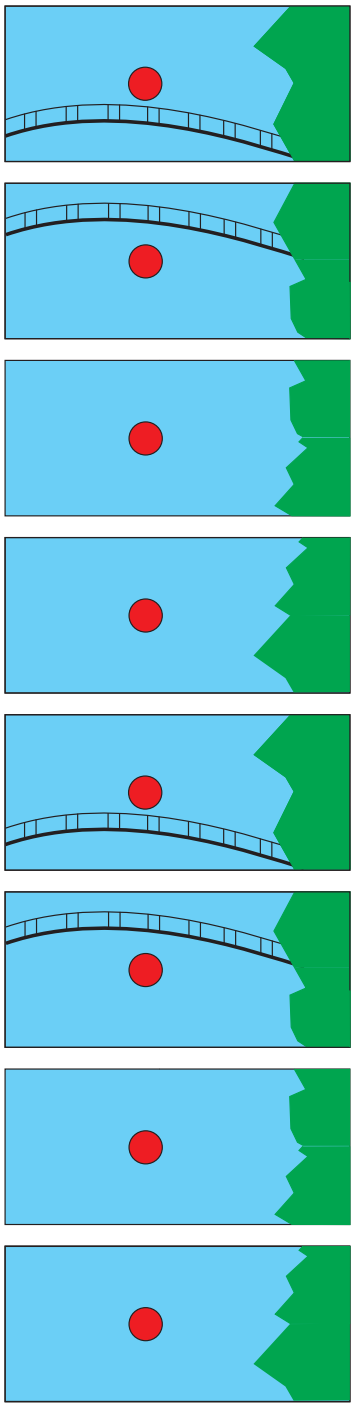

Figure 6. The changing backdrop for a falling object in a cartoon, illustrating the use of an abstract-iconic temporal text. The sequence of four frames, when repeated behind the cartoon object, gives the impression that the object is falling.

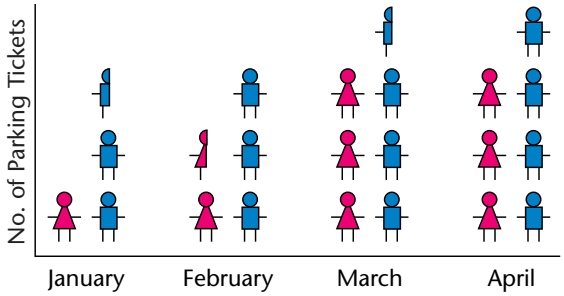

Figure 7. An iconic chart, illustrating the use of an abstract-iconic schematic text. Here, icons of men and women are used to represent numbers.

shows examples in the visual modality. Note that the syntax of augmentation focuses on the individual sign's shape, color, or font.

Aural modality

When applying this model to the aural modality, two additional issues need to be considered.

First, the difference between concrete-iconic and abstract-iconic texts is the difference between recordings of real sounds and sounds that have been artificially synthesized. This distinction thus proves less useful in the aural modality than in the visual modality.

Second, since by nature the aural modality is temporal, it is difficult to define individual objects of aural communication without considering the temporal dimension. Taking into account the duration and possible decomposition of the communication, I define individual and augmented aural objects as "very brief, atomic, aural texts, which communicate a single concept." Temporal aural texts also communicate a single concept, but their duration need not be brief, and in linear aural texts, the message may change over time.

Table 2. Aural modality examples.

	Concrete-Iconic	Abstract-Iconic	Symbolic
Individual	A recording of a brief, atomic sound, such as a car ignition	A brief, atomic synthesized sound, such as a "whirr" from a computer	A brief, atomic, symbolic sound like a doorbell
Augmentation	A recording of a brief, atomic sound, whose volume is significant, such as a door slammed in anger	A brief, atomic synthesized sound whose tone is significant, such as a desktop trash can that produces a "clunk" that decreases in tone as it fills	A brief, atomic, symbolic sound whose tone is significant, like an error "beep" that changes in tone according to the nature of the error
Temporal	A continuous recording representing a single concept, such as the sound of waves on a beach	A continuous synthesized sound representing a single concept, such as the sound of gunfire in a violent arcade game	A continuous symbolic sound, such as a fire alarm
Linear	A sequential recording of sounds representing a story, such as the build-up, height, and conclusion of a storm	A sequence of synthesized sounds, such as a train's approach, passing, and department that has been synthesized rather than recorded	A sequence of symbolic sounds, such as a computer "humm" that changes in pitch depending on the load on the network
Schematic	A recording of a sound comprising different frequencies, such as a car crash involving breaking glass and severe body damage	A synthesized sound track for an animated cartoon comprising different frequencies, such as a cat howling as it hits a solid wall	A complex sound where differing frequencies have different interpretations, such as a two-tone fire alarm indicating both location and severity of the fire

Table 2 shows examples in the aural modality. Note that in the aural modality, the syntax of augmentation focuses predominantly on the individual sign's tone, amplitude, or timbre. Also, the schematic arrangement in the aural modality is defined with respect to the frequency domain; thus, the second dimension of spatial aural communication is not space (as in the visual modality), but frequency. It's not easy for humans to distinguish the many different individual frequencies in a complex aural text like speech, and the examples are limited to aural texts with only two perceivable different frequency bands.

Composite-texts

Texts are either single-texts or composite-texts. Single-texts use only one modality and consist of one or more signs arranged according to a single syntax. They therefore embody a single representational system.

The definition of a composite-text is recursive. A composite-text contains more than one component-text, where the component-texts may be either single-texts or composite-texts. Thus a composite-text may embody more than one representational system and may use more than one modality.

The component-texts must themselves be arranged according to a syntax within the composite-text. The individual and augmentation syntactical arrangements are inappropriate for arranging more than one object. A composite-text will therefore have a temporal, linear, or schematic syntax arrangement associated with it.

For example, an instructional video (a composite-text that uses a linear syntax) may include the following component-texts:

- film of a lecturer explaining a problem (concrete-iconic, linear, visual),
- the soundtrack for the film of the lecturer explaining the problem (symbolic, linear, aural),
- some photographs (concrete-iconic, individual, visual), and
- some written paragraphs (symbolic, linear, visual).

This concept of composite-texts ensures that augmentation (which applies only to individual objects) may now be applied more generally over an entire text.

Extending the syntax dimension: the network category

The final syntactic category—network—does not describe the syntax for individual objects. It's only used for the arrangement of component-texts within a composite-text. Unlike the first five

Table 3. Visual and aural modality examples in the network arrangement.

	Concrete-Iconic	Abstract-Iconic	Symbolic
Visual modality	Interactive video, such as a video story where the reader chooses the story line	Interactive animation, such as an animated version of a video story	Hypertext, as an online thesaurus with links between related entries
Aural modality	Interactive audio of concrete-iconic audio recordings, such as a collection of different bird songs that can be selected individually	Interactive audio of abstract-iconic sounds, such as a collection of synthesized sound effects that can be selected individually	Interactive audio of symbolic sounds, such as touch-tone menus of recorded spoken information

syntactic categories, it does not restrict the order in which the user receives the component-texts.

In the network arrangement, component-texts connect together in a network structure of nodes and links, with related component-texts linked to each other. There is an implicit lack of linearity and no predefined sequence of receiving the entire composite-text. Thus, in receiving the text, a component-text may be followed by any one of the other component-texts associated with it. (Note that the network syntax corresponds to the hyper- prefix used in the terms hypertext and hypermedia.) Table 3 shows examples of the visual and aural modalities in the network arrangement.

The principle of synchronicity

When considering composite-texts that use more than one modality, it becomes possible (and indeed, sometimes essential) to transmit more than one message at once, as the receiver can now receive messages through each of the modalities used. The synchronous messages in the different modalities may be considered independent of each other from a perceptual point of view, although the receiver will usually make a cognitive link between them (for example, associating a "beep" from a computer with a visually perceptible error). Note that cognitive links may sometimes be made between two unintentionally perceptibly synchronous messages.

A synchronous-text, therefore, is a special type of composite-text containing an aural component-text and a visual component-text (which may be composite-texts themselves). The two component-texts are transmitted simultaneously, with the intention that the receiver make a cognitive link between the aural and visual perception.

Defining multimedia communication

This concrete and well-defined model of media objects serves as a fundamental basis for defining multimedia communication. The model itself does not suggest a single definition—it can be used to create different definitions of varying inclusiveness, as shown by the following examples.

1. The production, transmission, and interpretation of a composite-text, when at least two of the component-texts use different representational systems.

This definition of multimedia is very broad. Examples of multimedia communication under this definition include a wall poster that includes a photograph, some written paragraphs, and a map; a sentence that uses more than one font; and an audio novel with a single narrator and at least one sound effect.

The use of "two" in the definition arises from the simple principle "two is greater than one, and one component-text does not comprise a composite-text." Of course, the definition could adapt to any number (n) replacing "two," but this could cause problems in defining composite texts with $n - 1$ component-texts using different representational systems.

2. The production, transmission, and interpretation of a composite-text, where at least two of the component-texts use different representational systems in different modalities.

This definition extends the previous one, emphasizing a need to have more than one modality in the text, with a similar justification for the choice of the number "two." Examples include an audio tour of an art gallery; a television commercial with a "voice-over" (the voice of an unseen narrator); and a greeting card that plays a tune when opened.

3. The production, transmission, and interpretation of a composite-text, where the network syntax is used at least once.

This definition concentrates on the hyper- prefix: If a network syntax is used, and the receiver has some choice over the order in which some or all of the component-texts are transmitted, the text is considered multimedia. Note that the definition includes no restrictions on the nature of the component-texts: they may all be of exactly the same representational system. Examples include a book; telephone touch-tone menus of recorded spoken information; and a text-based interactive fiction story.

4. The production, transmission, and interpretation of a composite-text, where the network syntax is used at least once, and at least two of the component-texts use different representational systems.

Like the previous definition, this one emphasizes the hyper- aspect of the text, but in this case it imposes restrictions on the nature of the component-texts. Examples include a hypertext system that emits a "beep" whenever a link is traversed; a pictorial encyclopedia; and an indexed compact disk of a variety of different sound effects.

5. The production, transmission, and interpretation of a composite-text, where at least one of the component-texts is a synchronous-text.

This definition resembles the second one, in that the use of more than one modality is important. Examples include a picture of a waterfall, accompanied by the sound of running water; a film with a sound track; and an interactive film with a sound track.

The model's scope and potential

These definitions of multimedia communication are important both for what they include and for what they omit. The model permits defining multimedia communication in terms of the text's nature, with respect to a well-grounded semiotic basis and the representational systems the text uses.

The model does not provide for consideration of other general human communication issues, encompassed within the sample definitions' phrase "production, transmission and interpretation." The model makes no reference to the nature of this production, transmission, or interpretation, although each of these steps may have varying features.

For example, a cartoon strip may be hand drawn or computer generated, a piece of music may be transmitted via an audio recording or an orchestra, and a newspaper may be read from beginning to end, or selected articles may be chosen at will. All these considerations are external to the definition of the text's nature and can therefore be considered separately. The model does not tie definitions of multimedia to specific hardware or to production and interpretation methods.

This is the model's power rather than a limitation. It acknowledges that multimedia communication is a complex process that cannot be simply defined.

Thus, while the existing model produces much broader definitions than those usually employed (concentrating as they do on the nature of the text, and ignoring technology and methods), further defining the nature of the production, transmission, and interpretation permits further refining of these definitions.

In addition, models of text production, transmission, and interpretation may be related to this existing media model. Appropriate inter-model mappings may also be defined. For example, any multimedia text including a component-text using a network syntax requires an interactive device.[7]

Related work

Two theoretical classifications of representational systems for multimedia communication have already been proposed, both based on existing terminology that defines existing media types (graphics, beep, written sentence, sound).

The model presented here is comparable to that defined by Arens et al.[10] in two respects: the augmentation syntax relates to Arens' concept of signs (or information carriers) having "channels," and his definitions of complex exhibits and simple exhibits are similar to the definitions of composite-text and single-text presented here.

The taxonomy proposed by Heller and Martin[8] consists of two dimensions. The media expression dimension is related to a concrete-abstract continuum of expression: Elaboration media provide a "real world" depiction of the message (such as a photograph), Representation media allow for the communication of a more abbreviated, more "stylized" version of the original information (such as a map), and Abstraction media rely on metaphor and require that cultural context and experience be considered (for example, a well-known symbol). These three media expression categories are similar to the three categories of

sign defined here (concrete-iconic, abstract-iconic, and symbolic).

For the second dimension in their taxonomy, Heller and Martin use well-known existing media types: text, graphics, sound, and motion.

Lohse's visual categorization

Lohse et al.[11] categorize visual representations experimentally rather than by intuition or theoretical notions. Their experiments, based on 60 different visual items, resulted in defining 11 distinct categories. These categories tend to be defined in terms of the nature of the information contained within the visual item, rather than with respect to the representational system used in its creation. For example, in some cases, Lohse's categories of "tables" and "time charts" may relate to the linear visual category defined here, but this depends on the nature of the information represented, not on its representational system.

Lohse's categories "photorealistic pictures" and "icons" encompass the three categories of individual visual texts defined here. His definition of "icons" includes both abstract-iconic and symbolic signs.

A difference of perspective

All three of these classifications are driven in some manner by knowledge of currently existing media types: The Arens classification is used to describe given, existing representational systems in terms of the stated characteristics. Both Heller and Martin, and Lohse et al., use existing media types to label their categories.

The broader method of classifying representational systems presented in this article defines underlying characteristics over any number of appropriate dimensions. It then identifies those existing representational systems that relate to each cell in the model produced. Working from the dimensions suggested by theoretical semiotic characteristics of representational systems to the practical instantiations of the media thus defined allows considering a broader spectrum of media. This ensures that the taxonomy is complete—it still includes semiotic systems without a given, defining name in the common parlance of multimedia studies.

Conclusion

The model presented here avoids the temptation of using labels associated with existing text types. Instead it considers the basic components of the text and how they are arranged.

Consequently, rather than defining multimedia communication in terms of methods or technology, many alternate definitions may be proposed, each clearly related to the representational systems and syntax the text employs. This approach encourages a very broad view of the different possible types of multimedia communication. It also provides a basis from which to consider the other aspects of the communication—production, transmission, and interpretation. **MM**

Acknowledgments

I am grateful to Robert Colomb and Glenn Sweeney for their contribution to details of the examples used in this article, to Rachelle Heller and Dianne Martin for their feedback on a preliminary presentation of this work, and to the Human Communication Research Centre, University of Edinburgh, where much of this work was performed.

References

1. L. Barfield, *The User Interface: Concepts and Design*, Addison-Wesley, Reading, Mass., 1993.
2. B. Shneiderman, *Designing the User Interface,* 2nd edition, Addison-Wesley, Reading, Mass., 1992.
3. J. Preece, *Human-Computer Interaction,* Addison-Wesley, Reading, Mass., 1994.
4. C.S. Peirce, *Semiotic and Significs: The Correspondence between Charles S. Peirce and Victoria Lady Welby*, C.S. Hardwick, ed., Indiana University Press, Bloomington, Ind., 1977.
5. J.S. Bruner, "On Cognitive Growth," in *Studies in Cognitive Growth*, J.S. Bruner, R.R. Olver, and P.M. Greenfield, eds., Wiley, New York, 1966, p. 1-67.
6. R. Arnheim, *Visual Thinking*, University of California Press, Berkeley, Calif., 1969.
7. H.C. Purchase, "A Semiotic Definition of Multimedia Communication," *Semiotica*, 1998 (to appear).
8. R.S. Heller and C.D. Martin, "A Media Taxonomy," *IEEE Multimedia,* Vol. 2, No. 4, Winter 1995, pp. 36-45.
9. R.M. Baecker et al., eds., *Readings in Human-Computer Interation*, Morgan Kaufmann,San Francisco, 1995.
10. Y. Arens, E. Hovy, and M. Vossers, "On the Knowledge Underlying Multimedia Presentations," in *Intelligent Multimedia Interfaces*, M.T. Maybury, ed., MIT Press, Cambridge, Mass., 1993, pp. 280-306.
11. G.L. Lohse et al., "A Classification of Visual Representations," *CACM*, Vol. 37, No. 12, Dec. 1994, pp. 36-49.

Contact Purchase at Dept. of Computer Science and Electrical Engineering, The University of Queensland, St. Lucia 4072, Australia, e-mail hcp@csee.uq.edu.au.