Adapting Multimedia Design to Context

A design framework for interactive, user context-adaptive, multimodal learning environments.

Jennifer Brungart,[*] Hari Sundaram,[**] Harini Sridharan,[***] Ankur Mani[****]
and David Birchfield[*****]

[*] *School of Design, Arizona State University*
USA, jennifer.brungart@asu.edu.
[**] *Arts, Media and Engineering Program, Arizona State University*
USA, hari.sundaram@asu.edu.
[***] *Arts, Media and Engineering Program, Arizona State University*
USA, harini.sridharan@asu.edu.
[****] *Arts, Media and Engineering Program, Arizona State University*
USA, ankur.mani@asu.edu.
[*****] *Arts, Media and Engineering Program, Arizona State University*
USA, david.birchfield@asu.edu.

Abstract: This paper involves the description and discussion of a proposed dynamic presentation scheme for learning environments. The presentation scheme is an investigation into how high school-level content (i.e. geography) might address and adapt to the comprehension level, existing knowledge base and worldview of individual students, plus be comprehended quickly and clearly. The spatio-temporal arrangement of content that is employed within the scheme is meant to respond to the relative preparedness of students to comprehend the content, as well as to respond to the nature and complexity of the joint comprehension of multimodal information. The form of the scheme may be described as functioning similarly to a "chapter" within a media-rich, electronic "textbook." This richness is dependant upon the simultaneous presentation of associated textual, visual and auditory information.

The relationships explored and addressed within the context models described in this paper include the semantic interrelationships among concepts, linguistic and statistical relationships and common sense rules. Also, the environment offers students some opportunity to process information (i.e. "the causes and effects of population density") in a thorough, holistic and active manner, through the analysis and synthesis of associated sets of content.

The abstract framework of the learning environment consists of a content repository, an intelligent interaction interface and a dynamic presentation engine. This framework is based upon what have been determined to be the necessary components of an experiential learning environment: the efficient representation of content; the opportunity for an optimal selection of information; context-sensitive and dynamic presentation synthesis; and temporal content adaptation.

A model mechanism was developed and implemented for the objective evaluation of user experience with–as well as for the comprehension of content presented within–the proposed dynamic presentation scheme. A pilot study evaluation of the model elicited some encouraging results. It seems that a multimodal user context may be a valid approach after which to model optimal dynamic user context for electronic media-based applications. For one, such a context may be applied to the temporal adaptation of information in order to create information summaries.

Key words: Experiential Learning Environment, User Context Adaptation, Interaction, Dynamic Presentation Scheme, Multimodal Information

1. Introduction

The proposed learning environment discussed within this paper is as an investigation into how high school-level, geography-related content (i.e. the causes and effects of population density) might be presented to adapt to the comprehension level and existing knowledge base of individual students, and be learned quickly, clearly and in a manner that prompts active cognitive processing. This learning environment provides students with opportunities for dynamic interaction with multimodal information; includes efficient spatio-temporal arrangements of information; adapts user context to content; adapts the presentation of content to the level of comprehension of individual users; and addresses preferences of active learners.

Intended as a learning environment and not as a reference source, this dynamic presentation scheme functions similarly to a "chapter" within a media-rich, electronic "textbook." The purpose of the environment is to support the comprehension of geography-related content through both reflective and active cognitive means, via a user context that is consistent with individual students' past interactions with the environment.

The contextual relationships explored within the presentation scheme include: semantic inter-relationships among concepts; linguistic and statistical relationships; and common sense rules. The spatio-temporal arrangements of information are based upon–and are responsive to–the relative preparedness of individual students to comprehend particular content. These arrangements also support joint comprehension of multimodal (i.e. visual, textual, auditory) content and respond to the informational complexity of individual media elements.

The abstract framework of the learning environment consists of a content repository, an intelligent interaction interface and a dynamic presentation engine. This framework features critical components of an experiential learning environment, such as efficient representation of content, opportunity for the optimal selection of information, context-sensitive and dynamic presentation synthesis, and temporal content adaptation.

A model mechanism was developed and implemented, in the form of a pilot study, for the objective evaluation of the quality of user experience with, and comprehension of content presented within, the learning environment.

2. Method

2.1 Subject

According to [1], the tradition of high school-level classroom learning materials among Western cultures has tended to benefit those learners with reflective cognitive preferences, who prefer to process information by thinking about observations. Unfortunately, this approach to learning is not compatible with the dominant preferences of active learners–approximately half of all people–who prefer to learn by making and doing (i.e.

using information in order to understand it). Educational content, within both print and electronic resources, is most often structured sequentially, emphasizes the classification of concepts and focuses on the representation of theory in the absence of actual experience with concrete information. High school-level students were selected as the target audience for this learning environment since these students, more than others throughout grades K-12, are denied regular educational experiences that extend beyond the presentation of abstract content in a linear manner.

Geography-related content–specifically, "the causes and effects of population density"–was selected for presentation within the learning environment. There is some information related to the causes and effects of population density that can only be fully understood through the comprehension of sensory-based information (e.g. proximity to water, noise pollution, etc.), the presentation of which–within an electronic environment–seemed to offer an interesting challenge.

Content was created specifically for and tailored to the objectives of the learning environment. Voluntary content standards for twelve disciplines taught in K-12 education were developed in recent years, under the leadership of discipline-based educational associations in the United States [2]. These standards emphasize the most important concepts within the twelve disciplines–those that are considered worthy of "enduring" understanding [3]. The standards for geography were referenced for the development of content presented within this learning environment (e.g. "Understands the nature, distribution and migration of human populations on the Earth's surface," "Understands how human actions modify the physical environment," etc.). The appropriation of content directly from existing, linear, reflection-based textbook resources would not have been appropriate for the development of an active, media-rich learning environment.

2.2 Materials

The richness of the learning experience demonstrated within this system is dependant upon the simultaneous presentation of visual, textual and auditory information. The capacity of the electronic environment to simultaneously convey multiple levels of information strongly suggests the potential of the electronic "textbook" as an educational tool. The information communicated through the three modes–image, text and sound–should be related and mutually supportive, but not repetitive. In other words, one media element should enhance, rather than illustrate, another.

Collages of both sound [4] and image [5] possess the capacity to evoke a sense of time and place, the details of an environment, etc. It is possible for one to make clear inferences based upon information communicated through either visual and/or auditory means. A "soundscape" that conveys a passing jumble of indistinguishable footsteps, voices, car engines and honking horns, indicates a busy urban street. In contrast, the sound of a light wind blowing through leaves, a single pair of footsteps on a gravel drive and a bird crowing in the distance suggests a rural locale. Additionally, the sensory-based details conveyed within these soundscapes, indicate the urban area as being more densely populated than the rural area. Details within images have the capacity to suggest the relative population density of a location, as well (e.g. the number of people in the image, the height of or distance between buildings, etc.).

Media elements that support a particular concept should be positioned in close proximity to offer a clear spatial indication of what information belongs together within the presentation scheme. Attention to this sort of detail is particularly important when content is arranged in a collaged (i.e. overlapped) manner, as it is within our system, in order to prevent users from accidentally pairing elements that are not intended to be directly associated with one another.

2.3 Procedure

The interface of the learning environment is screen-based and consists of three separate maps (i.e. an urban location, suburban location and rural location) that represent associated sets of concepts to be explored and learned. Various meaningful locations on the maps, which correspond with population density-related content, are rollover-sensitive. When one rolls over a sensitive area, a set of relevant textual, visual and auditory information is revealed in a preset, timed sequence which forms a collage of text, image and sound–the text and images of which eventually nearly fill the screen (See Figure 4).

The learning environment is structured in a way that allows students some opportunity for choosing the order in which the categories of content (i.e. "urban," "suburban," "rural") may be accessed. The flexibility afforded in allowing students to make choices in how to pursue learning is of particular importance to active learners in their achievement of cognitive satisfaction [1, 6].

Keywords relevant to the concepts appear within a dedicated section at the right side of the interface. This is to allow students to quickly and easily refer to the terminology that is important for them to know and understand. Concrete words possess the potential to add precision to communication and may be used to cast unfamiliar information with familiar details [7]. This method of relating new information to existing knowledge may help to establish the information as relevant and worthy of attention. According to [8], as a function of preservation, the human brain is wired to dismiss information that it does not perceive to be relevant.

The manner in which content is presented within this learning environment requires that, initially, the information be processed in a reflective manner as it appears on the screen. However, it seems that the cognitive process that a student must undertake in order to analyze and synthesize the new information into her own holistic understanding of the causes and effects of population density is, indeed, quite active. Analysis and synthesis are two domains of higher order thinking articulated within Bloom's Taxonomy [9] and the ability to analyze and synthesize information is highly valued in today's world [7].

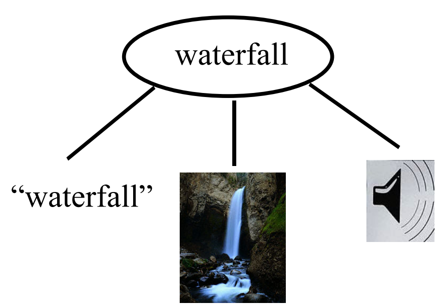

Figure 1: The concept "waterfall" is represented with text, image and sound.

2.3.1 Context Model

The context model for our presentation scheme is defined using a semantic-net – a graph $G = <V,E,W>$ where the nodes $v_i \in V$ represent the concepts, the edges $e_{ij} \in E$ represent the type of relationship (e.g. semantic, spatio-temporal) between the nodes i and j, and $w_{ij} \in$ W, specifies the strength of the relationship between the nodes (see Figure 2) [10, 11]. A concept node may be associated with an image, video, audio segment, text or combination of one or more of these.

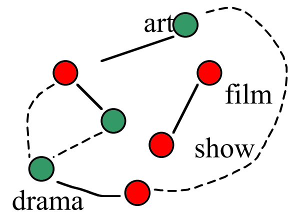

Figure 2: A semantic net.

2.3.2 User context

We define the user context to be comprised of concept nets composed of: the initial user profile (i.e. user interests, background, etc.); the viewing history (i.e. knowledge acquired while browsing the environment); user behavior (e.g. locations visited and time spent per media element); and user learning goals (see Figure 3).

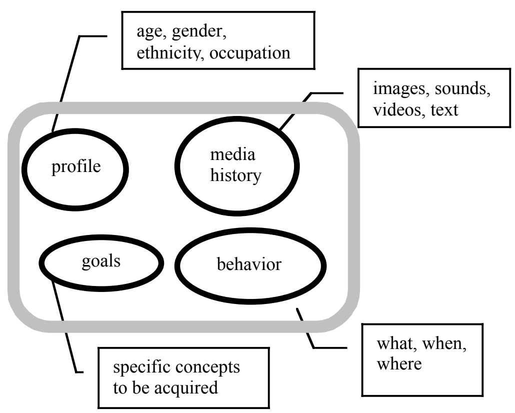

Figure 3: The user context model has four components:
profile; media history; behavior; and goals.

2.3.3 Context evolution

Our earlier framework [11] has been extended to incorporate multiple media (i.e. text, audio, images). Thus, the initial input of users' profiles may be in the form of text, audio clips and/or images.

The user context is referenced for the selection of media to present to the user. As the user interacts with the environment, the user context changes based upon the media consumed and time spent on each media element.

We used a leaky bucket model from [11] as the basis for the evolution of context in our presentation scheme. Concepts automatically acquired by the user profile through the system are slowly lost over time at a fixed rate, unless the user reinforces the concepts by visiting related media. The process of context evolution here is analogous with human memory, which is an essential part of experience that prompts mental associations. The reinforcement of concepts in our model involves the formation of associations between similar events.

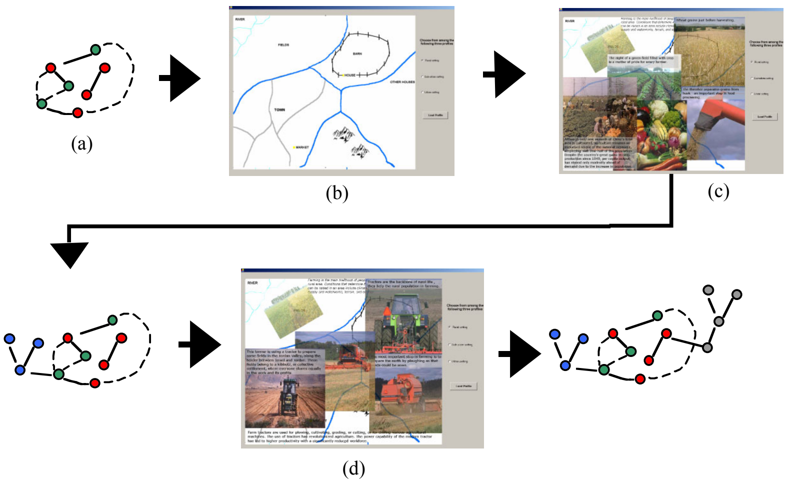

Figure 4: The graphs [(a), (c), (d)] represent the context of the same user at different times. The map (b) shows a specific location used by a student to learn about population density. The student does a mouse-over over the "fields" location on the map, and the system responds by dynamically creating a collage [see (c)] of images, text and sounds. As the user continues, the system refines the presentation of information as the user acquires additional information. The user context is updated to reflect the student's increased knowledge.

For each collage visited by the user, the experiential system creates a semantic net from the media and measures the time spent by the user on each collage. The interaction results in certain concepts being introduced into the user context (i.e. new knowledge), certain concepts being reinforced through association with similar concepts already in the user context, and certain concepts being disposed of over time. We have used the growth and decay equations given in [11] to calculate whether to reinforce or dispose of concepts based on time spent by the user with each event.

2.3.4. Optimal media selection

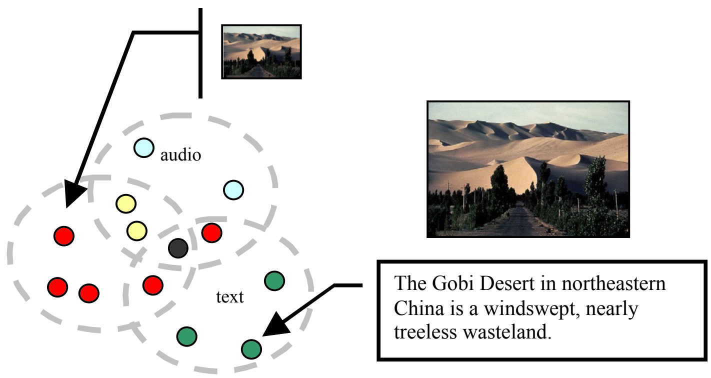

Figure 5: A disciplinary expert associates a set of images, sounds and text with each concept. (Each media element is represented as a full circle).

In our multimodal environment, individual concepts are represented through various media elements that interrelate and co-occur as media sets, both spatially and temporally. In the current system, this is done by disciplinary experts, although it could well be automated via media annotations.

Each concept may be represented by a large number of media sets, from which subsets of media elements are selected to maximize the coherence of information with respect to the user context and the location of the elements relative to the map. The subsets are chosen to complement the current user context, or distance from the user context (see **Error! Reference source not found.**), and to be relevant to the location selected on the map.

2.3.5. Dynamic presentation synthesis

The most important and distinguishing feature of this system is the dynamic presentation synthesis that seeks to optimize learning and coherence. For any given media set and presentation scheme (e.g. collage), there are various ways that the media elements can be organized, both spatially and temporally. For example, a collection of images presented within a collage could be arranged arbitrarily. Our presentation synthesis problem, however, examines how an *optimal* arrangement of images can be demonstrated within a collage and how the images might be displayed over an optimal amount of time.

2.3.6 Media presentation duration

Once a media set is selected to represent a particular concept, our system estimates the optimal time for presentation so that the media elements may be well comprehended. The effective duration required for the presentation of a media element is contingent upon the amount of time it takes a user to properly comprehend its information content (i.e. the complexity of the media element). For example, it is known that in films the close-ups and less complex shots seem to last longer than shots with much detail [12] and this idea has been successfully used to calculate comprehension time and generate the visual skims in [13]. Similar ideas exist for audio comprehension, as well, and experiments in cognitive neuro-psychology show that sentence comprehension is related to the structural complexity of the sentence [14]. Indeed, through our own experimentation, we have found that the joint comprehension time for a set of media elements (i.e. visual, audio and textual elements) depends upon the compressibility of the individual media elements.

2.3.7 Presentation order

The manner in which the presentation of information is ordered within the learning environment is crucial to the proper comprehension of key concepts. A good presentation scheme helps students learn concepts quickly, and supports freedom of interaction with the environment. Also, the coherence of information is an important factor in maintaining user attention and increasing learning capacity.

In our presentation scheme, the goal is for students to learn a set of concepts related to causes and effects of population density. Scores have been assigned to each media element that indicate the relative importance of the particular information to learning about population density. Additionally, these media elements have been divided into '*associated sets*' in support of particular concepts.

3. Results and Discussions

Table 1: User studies on our interactive environment

The environment is coherent.	6.2 / 7
The media concepts shown in increasing knowledge progression.	5.0 / 7
The speed of the presentation is just right.	7.0 / 7
In the environment, the text, audio and images appeared correlated.	6.8 / 7

We conducted experiments to evaluate the coherence and the effectiveness as a learning environment of our presentation system.

We evaluated our models through a pilot study that included five users. Two different presentation systems were created and used in the experiment: one with the dynamic presentation synthesis discussed previously and one with the media elements presented in a random manner. The users were asked to interact with the systems by freely navigating through different locations on the map interfaces and then comparing the two systems with respect to the coherence and comprehensibility of the concepts presented. All users found the system featuring dynamic presentation synthesis to be more coherent and comprehensible.

The users were also asked a set of questions geared toward the evaluation of the coherence and comprehensibility of the presentation and the correlation between different media elements. The users were asked to rate these aspects of the system on a scale of 1-7, with 1 representing "strongly disagree" and 7 representing "strongly agree." The results obtained may be seen in Table 1. Using the standard t-test, these results are statistically significant, at confidence value of 0.99 or better.

4. Conclusions

The user study evaluation of our investigation into how high school-level, geography-related content might be presented to adapt to the comprehension level and existing knowledge base of individual students elicited some encouraging results. For example, a multimodal user context is, indeed, a novel approach after which dynamic electronic media-based experiences may be modeled.

The development process and evaluation of this proposed learning environment has prompted many new questions, however, in regard to the very nature of user context as it relates to the comprehension of information. As an early attempt, there are, of course, refinements to be made and further studies to be undertaken with this dynamic presentation scheme.

It is possible that this work may evolve into the exploration and development of an adaptable, dynamic presentation scheme that seeks to facilitate and support opportunities for increased “making and doing.” For example, a next step may emphasize the development of opportunities for the creation of visual (e.g. sketches, maps, charts) and/or verbal (e.g. descriptions, stories, lists) representations of comprehension. Additionally, future exploration may include the implementation of content from other academic disciplines besides geography into the presentation scheme. This system is meant to serve as a model for adaptable, dynamic learning and not for the demonstration of content specific to a single discipline.

References

[1] McCarthy, B., 1987, The 4MAT system: teaching to learning styles with right/left mode techniques, Excel, Barrington, IL.

[2] Kendall, J.S. & Marzano, R.J., 1996, Content knowledge: a compendium of standards and benchmarks for K-12 education, Mid-continent Regional Educational Laboratory, Aurora, CO.

[3] Wiggins, G.P. & McTighe, J., 1998, Understanding by design, Association for Supervision and Curriculum Development, Alexandria, VA.

[4] McAdams, S. & Bigand, E., 1993, Thinking in sound: the cognitive psychology of human audition, Clarendon.

[5] Adams, J.L., 1986, Conceptual blockbusting: a guide to better ideas, 3rd Ed., Addison-Wesley, Reading, MA.

[6] Brosterman, N. & Togashi, K., 1997, Inventing kindergarten, H.N. Abrams, New York.

[7] Davis, M., 2002, Models and diagrams (draft), North Carolina State University, Raleigh, NC.

[8] Sperber, D. & Wilson, D., 1995, Relevance: communication and cognition, 2nd Ed., Blackwell, Cambridge, MA.

[9] Bloom, B.S., 1956, Taxonomy of educational objectives: the classification of educational goals, 1st Ed., Longmans Green, New York.

[10] Shevade, B. & Sundaram, H., 2003, Vidya: an experiential annotation system, paper presented at the 1st ACM Workshop on Experiential Telepresence, in conjunction with ACM Multimedia 2003, Berkeley, CA.

[11] Sridharan, H., Sundaram, H., Brungart, J. & Birchfield, D., 2004, Geography for active learners, paper presented at Arts , Media and Engineering Program , Arizona State University, AME-TR-2004-01, 2004.

[12] Sharff, S., 1982, The elements of cinema: toward a theory of cinesthetic impact, Columbia University Press, New York.

[13] Sundaram, H. & Chang, S.F., 2001, Constrained utility maximization for generating visual skims, paper presented at the IEEE Workshop on Content-Based Access of Image and Video Libraries, 2001, Kauai, HI.

[14] Caplan, D. & Waters, G., 1999, Verbal working memory and sentence comprehension, Behavioral and brain sciences, 22, 77-126.