Experiential Media Systems

HARI SUNDARAM, Arizona State University

This paper presents a personalized narrative on the early discussions within the Multimedia community and the subsequent research on experiential media systems. I discuss two different research initiatives—design of real-time, immersive multimedia feedback environments for stroke rehabilitation; exploratory environments for events that exploited the user's ability to make connections. I discuss the issue of foundations: the question of multi-sensory integration and superadditivity; the need for identification of "first-class" Multimedia problems; expanding the scope of Multimedia research.

Categories and Subject Descriptors: H.1.2 [**MODELS AND PRINCIPLES**]: User/Machine Systems—*Human information processing*

General Terms: Human Factors

Additional Key Words and Phrases: Multimedia, Experiential Systems, Foundations

ACM Reference Format:

Hari Sundaram. 2013. Experiential Media Systems *ACM Trans. Multimedia Comput. Commun. Appl.* 2, 3, Article 1 (May 2013), 4 pages.

DOI: http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Our encounters with the world are rich multi-sensory experiences. As you are reading this document, you are aware of the sounds around you, the touch of your fingers on the computing device, or perhaps the feel of paper if you are reading it in print. Consider the act of swimming, as another example, where, as you are propelled forward by the movements of your hands and legs, water rushes around you, providing you detailed multi-sensory feedback (touch, sight, sound) on your actions. In each of these examples, our multi-sensory experience of the world appears seamless and in harmony.

Could we construct artificial, real-time, multi-sensory environments that captured the "richness" (engaging, embodied, holistic) of natural experiences, where individuals could use their entire body to interact with the environment? Why were discussions about holistic, embodied experiences, seemingly natural to the field of Multimedia, absent from its research discourse? These questions on the idea of "experience" were my mind in the spring of 2002, while I was completing my doctoral work on summarizing audio-visual sequences [Sundaram 2002]. Furthermore, I worried that my dissertation wasn't contributing to "foundational" multimedia theory. Unlike fields of Computer Vision or Audition, whose core ideas were tied to how human beings perceived light or sound, Multimedia appeared to lack

Author's address: H. Sundaram, Room 398, 699 S. Mill Ave, School of Arts, Media and Engineering, Tempe, AZ 8528; email: hari.sundaram@asu.edu

DOI: http://dx.doi.org/10.1145/0000000.0000000

a connection to how we experienced our world. When I expressed some of my concerns to Larry Rowe, the outgoing SIGMM chair, at the 2002 ACM Multimedia conference held at Juan-les-Pins, he counseled for patience; he pointed out that the early years in the Database community resembled the Multimedia community in its lack of a coherent voice.

In this paper, I present a personalized narrative of experiential media systems research, including origins and definitions and two different research initiatives. In Section 3, I discuss reasons for absence of foundational multimedia theories, as well as the need to identify "first-class" Multimedia problems. I conclude with how Multimedia research can become the field that addresses complex societal challenges.

2. MULTIMEDIA RESEARCH AND THE IDEA OF EXPERIENCE

In the Fall of 2002, through a fortunate series of events, I found myself with an interdisciplinary faculty appointment between the arts and engineering at Arizona State University. A key realization as a faculty was that the absence of foundational multimedia theory was an opportunity for a young scholar—I wasn't bound by tradition, unlike peers in say Computer Vision, in my choice of problems and experimental methods. A conversation with Ramesh Jain, the incoming SIGMM chair, at the Juan-les-Pins Multimedia conference, provided tremendous support—not only was he keen to see new perspectives in the Multimedia community emerge, but also he was working on experiential computing.

My research on experiential environments was influenced by several disciplines—Robotics (Rodney Brooks, artificial life), AI (Stuart Russell, resource bounded reasoning), HCI (Paul Dourish, embodied interaction) and the Arts (Xenakis, stochastic frameworks for aesthetics). There were three principal insights: to make sense of the physical world, human beings actively use their body to interact with and alter their environment; we act under cognitive and physical constraints and employ approximations to reason about and act in our world; there is a fertile ground between New Media Art [Jordan and Packer 2002; http://www.w2vr.com] and Multimedia computing for the design of experiential systems.

These complementary perspectives led to a natural question: *How do we design immersive real-time multimedia feedback environments where human beings interact freely with their whole body and where the semantics of interaction are interpretable and exist at multiple levels?* In other words, can we replicate the richness of a real-world experience in an artificial environment? The answer to this question could enable new applications: mechanisms for rehabilitation; multimedia playgrounds to teach children; systems that enable sportspersons to improve their abilities (golf, tennis, swimming). Notice that in all of these examples, instead of only identifying patterns in data, the goal is to *transfer* semantics to the individual immersed in an environment. These are *machine teaching* problems, to contrast with the familiar learning problem.

Frank Nack, Gopal Pingali and I came together to organize the first SIGMM workshop to address the question of experiential multimedia systems [Sundaram et al. 2003]; the workshop was co-located with the 2003 ACM Multimedia conference in Berkeley. The workshop brought together researchers from many different perspectives ensuing in a lively debate on the notion of "experience." At the workshop Marc Davis [Davis 2003] made a careful distinction between "data" and "experience"—multimedia systems can only sense, model and transmit *data*, whereas it is the human mind that can occasion an *experience*. The workshop was also timely, as the first SIGMM strategic retreat held a few days prior to the Berkeley conference had identified experiential telepresence as one of the three grand challenges for the Multimedia community [Rowe and Jain 2005].

Over the next decade, Thanassis Rikakis and I began to collaborate on the design of a specific experiential multimedia system for stroke patient rehabilitation [Chen et al. 2006]. We defined experiential media systems to refer to real time, physically grounded multimedia systems where the participant was both the producer and consumer of meaning. These systems required embodied interaction on part of the user (that is, the participant used her body to interact with the system, moving beyond

the keyboard-mouse interaction metaphor) to gain new knowledge. We chose to work with stroke rehabilitation as our first experiential system, because the rehabilitation goal—to reach and grasp an object—resulted in unambiguous interaction semantics, simplifying the analysis of human activity and the design of multimedia feedback. Our system showed increased rate of learning of the rehabilitation goal for stroke survivors and was successfully deployed within a leading hospital in Phoenix, AZ.

Ramesh Jain [Jain 2003] adopted a different approach towards experiential computing. Instead of working on design of immersive multimedia environments, his research group focused on systems for supporting exploratory interaction with data. His group developed systems for understanding events from multiple perspectives that exploited the reasoning capabilities of the user navigating the data. Westermann and Jain [2007] broke new ground by suggesting that *media are event metadata*, instead the usual framing in both computer vision and multimedia, where we search for events *within* media. The paper framed interpretation and search of human events as a "first-class" multimedia problem, where heterogeneous data sources, aggregated at multiple time-scales are required to interpret events.

Experiential systems illustrate the overarching idea of *embodiment*: human beings use their body, with attendant cognitive and physical resource constraints, to make sense of the world.

3. REFLECTIONS ON MULTIMEDIA FOUNDATIONS

Let us return to the question asked in opening section: what is foundational about multimedia analysis? Why don't we have a multi-sensory "primal sketch?" There is a good reason why we know so little about the physiological basis of integration of the different senses. The early experiments in audition and vision, were never designed to understand the relationship between the senses—to prevent confounds, the experiments were designed to vary the signal in one modality only, while keeping the signals from the other modalities constant [Stein and Meredith 1993]. The positive news is that understanding the relationship between the senses is an emerging area of study [Stein et al. 2009].

The Multimedia community needs to identify and to focus its energies on "first-class" Multimedia problems, while we await a better understanding of sensory integration. I use the word "Multimedia" in the following sense: a Multimedia framework is one that employs sensing, and analysis, manipulation (including applications) and distribution of *multiple* data types, including audio, visual, social communication, and sensor data (GPS, accelerometer, biosignals). For "first-class" Multimedia problems, when we integrate heterogeneous data types, we should expect an outcome similar to "superadditivity" [Stanford and Stein 2007]: the effect of integrating multiple data types significantly exceeds the sum of the individual effects for each data types. The superadditivity effect is especially prominent in the integration of heterogeneous, low SNR, data streams. In other words, "first-class" Multimedia problems refer to the class of problems that can *only* be satisfactorily addressed by integrating heterogeneous data. Experiential media systems are examples of solutions to a category of "first-class" problems—one cannot reconstruct the holistic, embodied experience of the natural world with a single modality.

We need to become more inclusive in our understanding of "Multimedia," as well as our interpretation of "semantics." As a community, we have long focused on the analysis, systems and applications related to images, video and sound. The *field* of Multimedia should refer to *all* problems that require the use of heterogeneous data types for a satisfactory solution. In other words, the use of *single* well known modality—images, video or sound—to solve a difficult problem, is good Computer Science, but unimpressive Multimedia research. The problem of "semantics" is typically formulated with the goal of learning the relationship between multimedia data and labels. This implicit compact limits Multimedia research. Labels are less useful within experiential media systems, for example, as they fail to capture the complexity of the experience. Finally, we don't ask with equal vigor, a more basic question—from where do these meanings and patterns emerge? As a community, we have paid little attention to the social construction of semantics, their evolution, or that interpretation varies with time-scale.

4. CONCLUSIONS

In this paper, I presented the early discussions and subsequent research on experiential media systems, that were motivated by a simple question: *can we replicate the richness of a natural experience within an artificial environment?* I briefly discussed two different initiatives to experiential computing, one where researchers designed immersive multimedia environments for stroke rehabilitation and the other where researchers developed exploratory environments for events that exploited the user's ability to make connections. Finally, I discussed the question of foundations, including reasons for absence of a multi-sensory primal sketch, the need to focus on "first-class" Multimedia problems.

Multimedia research can carve out an important niche in Computer Science. To do so, we need to focus on "first-class" problems that truly require multiple data types for a satisfactory solution (e.g. problems that exhibit superadditivity), and by expanding the scope of Multimedia research beyond historically significant data types including images, video and sound. Experiential media systems are examples of solutions for "first-class" problems where Multimedia research can make a significant, unique contribution. Excitingly, for new generation of Multimedia researchers, there remain important unsolved problems in experiential media systems. It will be crucial for the field to emphasize problems that need heterogeneous data integration over problems that can be satisfactorily addressed with one data type. Addressing important societal challenges in public health, managing an aging population, and sustainability, for example, requires us to capture, interpret and distribute heterogeneous data types (including sounds, videos, sensor data, social communication) from a population. These challenges are non-trivial because they are data sparse, semantics arise due to interaction between data types and interpretation varies with time scale. By becoming the field that solves problems requiring heterogeneous data integration and at large scale, we distinguish ourselves from our peers—Computer Vision, Natural Language Processing and Audition—in addressing complex societal challenges.

REFERENCES

Yinpeng Chen, He Huang, Weiwei Xu, Richard Isaac Wallis, Hari Sundaram, Thanassis Rikakis, Todd Ingalls, Loren Olson, and Jiping He. 2006. The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation. In *Proceedings of the 14th annual ACM international conference on Multimedia*. ACM, 763–772.

Marc Davis. 2003. Theoretical foundations for experiential systems design. In *Proceedings of the 2003 ACM SIGMM workshop on Experiential telepresence*. ACM, 45–52.

Ramesh Jain. 2003. Experiential computing. *Commun. ACM* 46, 7 (2003), 48–55.

Ken Jordan and Randall Packer. 2002. *Multimedia: From Wagner to Virtual Reality*. W. W. Norton & Company.

Lawrence A. Rowe and Ramesh Jain. 2005. ACM SIGMM retreat report on future directions in multimedia research. *ACM Trans. Multimedia Comput. Commun. Appl.* 1, 1 (Feb. 2005), 3–13. DOI:http://dx.doi.org/10.1145/1047936.1047938

Terrence R Stanford and Barry E Stein. 2007. Superadditivity in multisensory integration: putting the computation in context. *Neuroreport* 18, 8 (2007), 787–792.

Barry E Stein and M Alex Meredith. 1993. *The merging of the senses*. The MIT Press.

Barry E Stein, Terrence R Stanford, Ramnarayan Ramachandran, Thomas J Perrault Jr, and Benjamin A Rowland. 2009. Challenges in quantifying multisensory integration: alternative criteria, models, and inverse effectiveness. *Experimental Brain Research* 198, 2-3 (2009), 113–126.

Hari Sundaram. 2002. *Segmentation, structure detection and summarization of multimedia sequences*. Ph.D. Dissertation. Columbia University.

Hari Sundaram, Gopal Pingali, and Frank Nack (Eds.). 2003. *Proceedings of the 2003 ACM SIGMM workshop on Experiential Telepresence*. ACM. 127 pages.

Utz Westermann and Ramesh Jain. 2007. Toward a common event model for multimedia applications. *Multimedia, IEEE* 14, 1 (2007), 19–29.

Received May 2013; revised July 2013; accepted October 2013