

**Sto(ry)chastics:  
a Bayesian network architecture for combined  
user modeling, sensor fusion, and computational storytelling for  
interactive spaces**

by  
Flavia Sparacino

M.Eng., Electrical Engineering, Politecnico di Milano (1993)  
M.Eng., Mechanical Engineering, Ecole Centrale Paris (1993)  
S.M., Cognitive Sciences, Universite' Paris VI (1994)  
S.M., Media Arts and Sciences (1997)

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2002

© Massachusetts Institute of Technology 2001. All rights reserved.

Author

---

Flavia Sparacino  
Program in Media Arts and Sciences  
October 15 2001

Certified by

---

Kent Larson  
Principal Research Scientist  
School of Architecture and Planning  
Thesis Supervisor

Certified by

---

Glorianna Davenport  
Principal Research Scientist  
Media Laboratory  
Thesis Supervisor

Accepted by

---

Andrew B. Lippman  
Chair, Departmental Committee on Graduate Students  
Program in Media Arts and Sciences



# **Sto(ry)chastics: a Bayesian network architecture for combined user modeling, sensor fusion, and computational storytelling for interactive spaces**

by  
Flavia Sparacino

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning  
on October 15 2001, in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy

## **Abstract**

This thesis presents a mathematical framework for real-time sensor-driven stochastic modeling of story and user-story interaction, which I call *sto(ry)chastics*. Almost all sensor-driven interactive entertainment, art, and architecture installations today rely on one-to-one mappings between content and participant's actions to tell a story. These mappings chain small subsets of scripted content, and do not attempt to understand the public's intention or desires during interaction, and therefore are rigid, ad hoc, prone to error, and lack depth in communication of meaning and expressive power. *Sto(ry)chastics* uses graphical probabilistic modeling of story fragments and participant input, gathered from sensors, to tell a story to the user, as a function of people's estimated intentions and desires during interaction. Using a Bayesian network approach for combined modeling of users, sensors, and story, *sto(ry)chastics*, as opposed to traditional systems based on one-to-one mappings, is flexible, reconfigurable, adaptive, context-sensitive, robust, accessible, and able to explain its choices.

To illustrate *sto(ry)chastics*, this thesis describes the museum wearable, which orchestrates an audiovisual narration as a function of the visitor's interests and physical path in the museum. The museum wearable is a lightweight and small computer that people carry inside a shoulder pack. It offers an audiovisual augmentation of the surrounding environment using a small eye-piece display attached to conventional headphones. The wearable prototype described in this document relies on a custom-designed long-range infrared location-identification sensor to gather information on where and how long the visitor stops in the museum galleries. It uses this information as input to, or observations of, a (dynamic) Bayesian network, selected from a variety of possible models designed for this research. It then delivers an audiovisual narration to the visitor as a function of the estimated visitor type, and interactively in time and space.

The network has been tested and validated on observed visitor tracking data by parameter learning using the Expectation Maximization (EM) algorithm, and by performance analysis of the model with the learned parameters. Estimation of the visitor's preferences, in addition to the type, using additional sensors, and examples of sensor fusion, are provided in a simulated environment.

The main contribution of this research is to show that (dynamic) Bayesian networks are a powerful modeling technique to couple inputs to outputs for real-time sensor-driven multimedia audiovisual stories, such as those that are triggered by the body in motion in a sensor-instrumented interactive narrative space. The coarse and noisy sensor inputs are coupled to digital media outputs via a user model, and estimated probabilistically by a Bayesian network. Other contributions are: the design of the museum wearable application, the assembly and fashioning of a wearable computer, specifically conceived for museum use; the design and realization of a new long-range infrared location-identification sensor; the construction and testing of a variety of Bayesian networks for user-type and profile estimation; the extension of the previous Bayesian network for real-time story-segment selection and editing; model selection; model validation and parameter learning via the EM algorithm; and simulation of processing multiple sensor inputs with a Bayesian network for more robust estimation and more accurate user profiling.

Other possible applications of sto(ry)chastics extend to digital storytelling for a variety of interactive architectural spaces, art installations, or the theater stage.

Thesis Supervisor: Kent Larson  
Title: Principal Research Scientist  
School of Architecture and Planning

Thesis Supervisor: Glorianna Davenport  
Title: Principal Research Scientist  
Media Laboratory



# Doctoral Dissertation Committee

Thesis Supervisor

---

Kent Larson  
Principal Research Scientist  
School of Architecture and Planning

Thesis Supervisor

---

Glorianna Davenport  
Principal Research Scientist  
Media Laboratory

Thesis Reader

---

Walter Bender  
Senior Research Scientist  
Media Laboratory  
Executive Director

Thesis Reader

---

Neil Gerschenfeld  
Associate Professor of Media Arts and Sciences  
Program in Media Arts and Sciences



# Contents

## **1. Introduction**

- 1.1. Interactive Spaces and new forms of communication
- 1.2. Telling stories in interactive spaces
- 1.3. Sto(ry)chastics and the museum wearable
  - 1.3.1. Contribution
  - 1.3.2. Document Layout

## **2. Motivation and Related Work**

- 2.1. A Taxonomy of Interactive System Architectures
- 2.2. Motivation
- 2.3. Requirements for new authoring tools
- 2.4. Related Work

## **3. Application: the Museum Wearable**

- 3.1. Scenario
- 3.2. The Museum Wearable
- 3.3. Visitor Types
- 3.4. Experimentation Platform: the museum wearable at MIT Museum's *Robots and Beyond* exhibit
- 3.5. Annotations and observations of visitors' behavior
- 3.6. The Museum Wearable: demonstration prototype

## **4. Estimating the visitor's intentions with Bayesian Networks**

- 4.1. Bayesian Networks
  - 4.1.1. Inference in Bayesian Networks
  - 4.1.2. Learning in Bayesian Networks
  - 4.1.3. Dynamic Bayesian Networks
  - 4.1.4. Related modeling techniques: HMMs, CHMMs, Kalman Filters, and Markov Random Fields.
- 4.2. Modeling visitors' intentions at MIT's Robots and Beyond Exhibit
  - 4.2.1. Visitor Type model 1
  - 4.2.2. Visitor Type model 2
  - 4.2.3. Visitor Type model 3
  - 4.2.4. Visitor Type models 4a and 4b
  - 4.2.5. Model selection

## **5. Sto(ry)chastics: editing stories for different visitor types and profiles**

- 5.1. Content granularity and the knobs of a computational storytelling machine.
- 5.2. Content selection for different visitor types
- 5.3. Content selection for different visitor profiles
  - 5.3.1. Adding content
  - 5.3.2. Adding visitor types
  - 5.3.3. Adding sensors: sensor fusion in a simulated environment

## **6. Building the Wearable**

- 6.1. The wearable computer
- 6.2. The choice of sensors
- 6.3. Sensor design: the infrared location sensor
- 6.4. The software
- 6.5. The head mounted display

## **7. Results And Evaluation**

- 7.1. Model validation
  - 7.1.1. Learning from the data
  - 7.1.2. Labeling the data
- 7.2. Comparison with previous real-time sensor-driven content selection architectures

## **8. The Impact of the Museum Wearable on Exhibit Design**

## **9. Summary of Accomplishments and Future Directions**

## **Bibliography**

## **Appendix. Sto(ry)chastics for other Applications**

# Chapter 1

## Introduction

### 1.1. Interactive spaces and new forms of communication

Our society's modalities of communication are rapidly changing. Large panel displays and screens are being installed in many public spaces, ranging from open plazas, to shopping malls, to private houses, to theater stages, and museums. In parallel, wearable computers are transforming our technological landscape by reshaping the heavy, bulky desktop computer into a lightweight, portable device that is accessible to people at any time. This combination of large public and miniature personal digital displays offer unprecedented opportunities to merge the virtual and the real, the information landscape of the Internet with the urban landscape of the city, to transform digital animated media in storytellers, in public installations and through personal wearable technology. Computation and sensing is moving from computers and devices into the environment itself. The space around us is instrumented with sensors and displays, and it tends to reflect a diffused need to combine together the information space with our physical space.

In this new communication age, interactive space design involves three elements: human authors (space designers) who conceive strategies and methods to deliver appropriate information to the public interactively; computers and sensors that gather and process *in real time* information about the public's behavior in the instrumented space; and the audience, with its own needs to receive personalized content only when and where it is appropriate.

Technological progress and miniaturization has produced off the shelf processors (computers) which are fast, small, lightweight and reliable, resulting in the creation of new needs for personalized, reconfigurable, and flexible information for the public at one end, and new authoring tools for the space designer on the other end. These tools need to be able to take input from the audience and deliver a personalized story articulated not only over time but also over space. Specifically, the digital architect who wishes to reshape our surrounding space and body, and transform them into technology-augmented devices for information exchange and artistic expression needs: sensors that are reliable and robust, and (mathematical) modeling tools which allow the system to understand the public's intentions and coordinate a narration.

This thesis focuses on the museum as an example of interactive narrative space. It introduces Bayesian networks for real time sensor-driven storytelling, and demonstrates that they are a powerful tool to model the uncertainty in the sensor measurements, make informed guesses about people's intentions during interaction, encapsulate the storyteller's message, and orchestrate a complex audiovisual narration as a function of

these. I call such stochastic modeling of story and user-story interaction: *sto(ry)chastics*. Sto(ry)chastics has implications both for the human author (designer/curator) who is given a flexible modeling tool to organize, select, and deliver the story material, as well as for the audience, who receives personalized content only when and where it is appropriate.

## 1.2. Telling stories in interactive spaces

Before I explain how *sto(ry)chastics* provides a flexible, reconfigurable and robust authoring tool for the space designer as it tailor a personalized story for the audience, I outline in this paragraph my approach and definition of story.

I borrow my definition of story for this research from the work for Jerome Bruner. Bruner is an American psychologist and educator whose work on perception, learning, memory, and other aspects of cognition in young children has, along with the related work of Jean Piaget, influenced the American educational system. In his book “Acts of Meaning”, Bruner [Bruner, 1990] defines story as a process by which meanings are created and negotiated within a community (p. 11). For Bruner, narrative is used to construct meaning by relating the individual or constituent aspects of human behavior to the context or situation in which a behavior occurs. He opposes construction of meaning through narrative, which he calls “narrative thought”, to objective information communication through “paradigmatic”, scientific-like reasoning [also in: Bruner, 1986]. My view of story and interaction is based on this distinction and privileges communication as narrative rather than communication as transmission of a message or information. For Bruner, paradigmatic reasoning shares with scientific explanation the mode of inductivism. Through it one sees a world of objects which interact in regular patterns. Narrative thought by contrast, attempts to maintain a subjective perspective on the world it represents, incorporating aims and fears into the picture. It incorporates at the same time a knowledge of the world and the point of view which beholds it. While Bruner defines story as a social and situated construction of meaning, his definition of story is an open-minded one, which shows a willingness to construe knowledge and values from multiple perspectives without loss of commitment to one’s own values. It asks that we be accountable for how and what we know, but it does not insist that there is only one way of constructing meaning, or one right way (p.30). For Bruner, narrative is used by individuals to create meaning through its dramatic quality. Using Burke’s analysis of story [Burke, 1969] with its five characteristics of actor, action, a goal, a scene, and an instrument, plus trouble (p. 50), Bruner argues that narrative involves both a cultural convention and a deviation from it that is explicable in terms of an individual’s intentional state. People use narrative to schematize their experience and this is a process that is situated socially and depends upon language. In this respect he recalls his earlier work *Essays for the Left Hand* [Bruner, 1962], where he proposed the existence of a “library of scripts” which are available to members of our culture as repertoires of understanding. It is exactly by considering our commonsense understanding of a situation (an item in Bruner’s library of scripts) that we can model a set of expectations and possible responses to a story fragment. This modeling is flexible and not rigid, as the

terms of the people-story interaction are governed in my approach by the improvisational nature of human communication as described below.

Improvisation, in music or theater, or even dance, is an interaction process in which individuals have some creative freedom, but at the same time are influenced by the situation and by each other's actions. In most informal and formal situations alike, individuals' actions are not scripted, yet a coherent, meaningful interaction results. This type of interaction model is similar to the improvised dialogue of children's social play as described by Vygotsky [Vygotsky, 1990]. When children imagine riding a horse while riding a broom, or when they imagine playing the role of a captain by wearing their parent's clothes, they exercise their creativity as they reinterpret situations they have observed or learned. They construct a new reality which responds to their needs and curiosities.

Sawyer has conducted a series of studies on the structure of improvisational interaction in children's pretend play and in theater [Sawyer, 1997a, 1997b], as well as music. He has developed a semiotic theory of improvisational interaction and has extended his study to social encounters, as examples of improvisational interaction. My definition of interaction is grounded on Sawyer's investigation on improv theater and children's pretend play. It is a process, in which the participants construct an emergent narrative having some creative and imaginative freedom. At the same time they are influenced by the situation and their knowledge of typical situations, as well as each other's actions.

When the computer is the storyteller, the notion of story described above, needs to be simplified and parametrized to a few elements that the computer can manipulate. A story narrated by a computer may come in the form of an audio-visual narrative. In the specific case of this research, this narrative is experienced through a mobile headphones/eye-glass display system that the user carries with them. For the purpose of this research, a story can be simply viewed as an ordered sequence of small audio-visual segments. Each of these segments is a closed mini-story with full meaning (sequence), and is to the whole story, what a sentence is to a paragraph of a written text. What this definition borrows from the previous discussion, in a simplified way is: 1. From Bruner: a story always needs to be situated in context, therefore to narrate a situated story, the computational storyteller needs to have a model of its audience. 2. From Sawyer: story is an emergent process which is not given all at start, but it is the result of interaction between the museum visitor and the system: it evolves with the user's path in the museum galleries.

### **1.3. Sto(ry)chastics and the museum wearable**

Sto(ry)chastics is grounded on the hypothesis that in order to build engaging interactive entertainment systems, able to be expressive and convey meaning and depth of content, we cannot have complex centralized programs which simply read sensor inputs and map them to actions on the screen. Interactive storytelling with such one-to-one mappings leads to complicated control programs which have to do an accounting of all the available content, where it is located on the display, and what needs to happen when/if/unless. These systems rigidly define the interaction modality with the public, as a consequence of their internal architecture. They need to carefully list all the

combinatorics of all possible interactions and then introduce temporal or content-based constraints for the presentation. Having to plan an interactive storytelling piece according to this methodology can be a daunting task, and the technology in place seems to somehow complicate and slow down the creative process rather than enhance or expand it. Rather than directly mapping inputs to outputs, we need to endow digital content itself with the ability to “understand the user” and to produce an output based on the interpretation of the user’s intention in the narrative context.

Sto(ry)chastics uses a dynamic Bayesian network to model the sensors and allows the system to interpret the sensor data by taking into account the context and domain of interaction, represented by other nodes of the network. The interpretation of sensor data is robust in the sense that it is probabilistically weighted by the history of interaction of the participant as well as the nodes which represent context. Therefore noisy sensor data, triggered for example by external or unpredictable sources, is not likely to cause the system to produce a response which does not “make sense” to the user. For content selection and delivery, sto(ry)chastics allows the system to build a profile of the participant through time, and therefore can tailor content according to the participant’s estimated desires and interests. These features: robustness with respect to “misunderstandings” because of knowledge of context, and the ability to learn more about the user through time, produce a system which with further development can potentially, in the future, simulate an elementary conversation with a human participant.

Another advantage of the Bayesian network approach is that the designer or programmer of the interactive experience only needs to set a general and not detailed structure of the story and user-story interaction. Bayesian networks can be trained from data and learn the right parameters for sequencing and interaction. This matches the improvisational nature of the interaction the designer needs to model: the designer only gives the system the general structure which describes its functioning. Then the parameters of the system are trained through a learning procedure and can be fine tuned in the course of interaction. This is new in the field of multimedia authoring: rather than giving the program all interaction and sequencing parameters at start, the designer infuses in the system only a general structure of what it should do. Then he/she trains the system to the task, basically by saying: “observe what I do and learn”. What happens then during interaction is similar to what happens during a musical jam session, in which the musicians follow a general well known set of rules, and yet they create a new piece which applies and modulates those rules according to the creative input of their imagination right there and then. This general structure, given to the program at start and mathematically described by the Bayesian network, models the library of scripts of our commonsense understanding of typical situations. Therefore using sto(ry)chastics designers will possibly be able to not only to create interactive environments which are compelling and robust but which can also interpret the user’s actions in context. The same action of the user can cause a different outcome according to the time slice in which that action is registered in the Bayesian network model.

Rather than describing sto(ry)chastics in general, it is easier to focus on a specific application, and use it as an example of modeling story and user-story interaction with



Bayesian networks. As an example of application of sto(ry)chastics, and to illustrate its features, I have designed and developed a real time storytelling device: a museum guide which in real time evaluates the visitor's preferences by observing his/her path and length of stops along the museum's exhibit space, and selects content from a set of available movie clips, audio, and animations. This device, which I call the Museum Wearable, illustrates the advantages of sto(ry)chastics in designing and authoring real-time sensor-driven digital media presentation systems. In this document I ground further discussion on modeling first the user's interest profile, and subsequently the selection of content, on this specific application.

### **1.3.1. Contribution**

The main contribution of this thesis is to show that (dynamic) Bayesian networks are a powerful modeling technique to couple inputs to outputs for real time sensor-driven multimedia narratives, such those that are triggered by the body in motion in a sensor-instrumented interactive narrative space. The coupling is done by interpreting sensor data as evidence which identifies the user's preferences, with probabilistic weights given by the context domain, modeled by appropriate nodes in the network, or by learning the parameters of the model. This approach can be considered robust, context-sensitive, flexible, reconfigurable, and extensible. Other contributions are: the project and design of the museum wearable application, the assembly and fashioning of a wearable computer, specifically conceived for museum use; the design and realization of a new long range infrared location identification sensor; the construction and test of a variety of Bayesian networks for user type and profile estimation; the extension of the previous Bayesian network for real time story segment selection and sequencing; model selection; model validation and parameter learning via the EM algorithm; and simulation of processing multiple sensor inputs with a Bayesian network for robust estimation and more accurate user profiling.

### **1.3.2. Document Layout**

This document is organized as follows:

- Chapter 2 describes a variety of techniques used to model interactive multimedia and highlights their advantages/disadvantages. It introduces the necessity for probabilistic modeling of sensors and content selection.
- Chapter 3 presents the museum wearable which is the application chosen and developed to demonstrate the advantages of sto(ry)chastics.
- Chapter 4 offers a short tutorial on Bayesian networks, and describes a variety of networks which estimate the visitor's type using only a location sensor.
- Chapter 5 illustrates sto(ry)chastics, and explains how it is applied to deliver a personalized story to the visitor with the museum wearable.

- Chapter 6 describes the hardware and software assembled and created for this research. It includes designs for the infrared location sensor custom made for the museum wearable.
- Chapter 7 shows how the Bayesian network developed for sto(ry)chastics was validated from real data and describes parameter learning from visitor tracking data gathered at the museum. It also offers an evaluation of the accomplishments of this research by comparison with other authoring techniques.
- Chapter 8 illustrates the potential impact of the museum wearable on exhibit and space design.
- Chapter 9 summarizes accomplishments and future directions of this research.
- The appendix outlines how to author other applications with sto(ry)chastics.

## Chapter 2

# Motivation and Related Work

### 2.1. A Taxonomy of Interactive System Architectures

Most of the existing interactive media applications can be categorized to be either scripted, or responsive, or occasionally, behavior-based [Sparacino, 2000b]. All of these systems imply a fixed modality of user-story interaction and story authoring. In this chapter I briefly describe advantages and disadvantages of traditional interactive story authoring system and illustrate the features that new authoring systems need to have to overcome most of the limitations of these previous authoring tools. In the following section I also describe in more detail examples of applications I developed as part of my PhD research, each from a different category of the previous taxonomy, which have all led to developing sto(ry)chastics as a solution for the limitations and problems encountered for each of them.

*Scripted systems* are those in which a central program coordinates the presentation of visual or audio material to the audience. The interaction modality is often restricted to clicking on a static interface which triggers new material to be shown. These systems need careful planning of the sequence of interactions with the public and acquire high complexity when drawing content from a large database. This authoring complexity often limits the experience to a shallow depth of content and a rigid interaction modality. Examples of scripted authoring technique can be found in [Sawhney, 1997].

*Responsive systems* are those in which control is distributed over the component modules of the system. As opposed to the previous architectures, these systems are defined by a series of couplings between user input and system responses. The architecture keeps no memory of past interactions, at least explicitly, and is event-driven. Many sensor-based real-time interactive art applications are modeled according to this approach. One-to-one mappings define a geography of responses whose collection shapes the system architecture as well as the public's experience. Although somewhat easier to author, responsive experiences are sometimes repetitive: the same action of the participant always produces the same response by the system. The public still tends to adopt an exploratory strategy when interacting with responsive systems, and after having tried all the interface options provided, is often not attracted to continue exploring the interactive experience. Sometimes simple responsive experiences are successful because they provide the participant with a clear understanding of how their input – gestures,

posture, motion, voice – determines the response of the system. The prompt timing of the response is a critical factor to be able to engage the public in the experience. Examples of responsive systems are described in [Davenport, 2000] and [Paradiso, 1999].

*Behavioral* systems or environments are those in which the response of the system is a function of the sensory input as well as its own internal state. The internal state is essentially a set of weights on the goals and motivations of the behavioral agent. The values of these weights determines the actual behavior of the agent. Behavioral systems provide a one-to-many type of mapping between the public's input and the system's response. The response to a particular sensor measurement or input is not always the same: it varies according to the context of the interaction which affects the agent's internal state. Successful behavioral systems are those which allow the public to develop an understanding of the causal relationships between their input and the agent's behavior. Theoretically, the public should be able to describe the dynamics of the encounter with a synthetic behavioral agent as they would tell a story about a short interaction with a living entity, human or animal. This is one of the reasons why behavioral agents are often called life-like creatures [Perlin, 1996], [Blumberg, 1995].

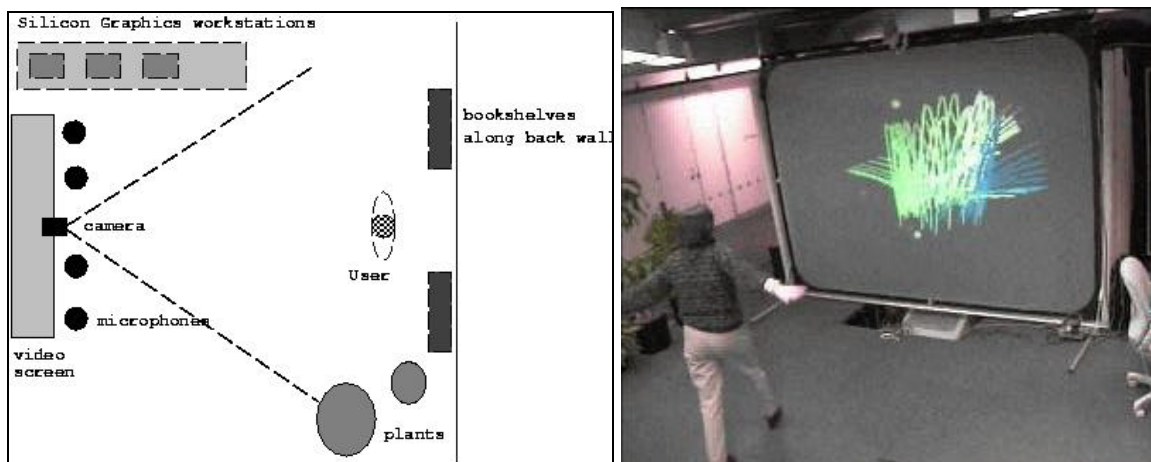
The behavior-based approach has proven to be successful when applied to mobile robots and to real-time animation of articulated synthetic creatures. In this context, "behavior" is given a narrow interpretation derived from behavioral psychology (Skinner). For animats, behavior is a stimulus-response association, and the action-selection mechanism which assigns weights to the layered behaviors can be seen as a result of operant conditioning [Catania, 1988] on the creature. Behavior-based AI has often been criticized for being "reflex-based", as it controls navigation and task execution through short control loops between perception and action. In my view, Skinner's reductive notion of behavior is insufficient to model many real life human interactions or simulated interactions through the computer. Multimedia, entertainment, and interactive art applications, all deal with an articulated transmission of a message, emotions, and encounters, rather than navigation and task execution. As we model human interaction through computer-based media we need to be able to interpret people's gestures, movements, and voice, not simply as commands to virtual creatures but as cues which regulate the dynamics of an encounter, or the elements of a conversation.

The previous taxonomy does not pretend to be exhaustive. It provides however a focus in defining a set of basic requirements, features, and architectures of current interactive media applications. A compelling interactive computer-based storyteller needs to have the depth of content of a scripted system, the flexibility of a responsive system, and the autonomous decentralized architecture of a behavioral system. It also needs to go beyond the behavioral scheme and respond not just by weighing stimuli and internal goals of its characters, but also by understanding the user's intentions in context, and by learning from the user.

## 2.2. Motivation

Most of my previous work uses an IVE (Interactive Virtual Environment) setup [Wren, 1997a]. IVE is an interactive space developed at the MIT Media Lab, in which the public interacts with visual material presented on a large projection screen which occupies one side of the room [figures 1,2]. A downward pointing wide-angle video camera mounted on top of the screen allows the IVE system to track a member of the public. By use of real-time computer vision techniques [Wren, 1997b][Darrell, 1996][Oliver, 1997] we are able to interpret the user's posture, gestures, identity, and movement. A phased array microphone is mounted above the display screen for audio pickup and speech processing. A narrow-angle camera housed on a pan-tilt head is also available for fine visual sensing. The only constraints are a constant lighting and an unmoving background.

IVE was built to enable people to participate in immersive interactive experiences without wearing suits, head-mounted displays, gloves, or other gear. Remote sensing via cameras and microphones allows people to interact naturally and spontaneously with the material shown on the large projection screen. IVE currently supports one active person in the space and many observers on the side. The IVE environment was originally developed for the ALIVE project [Darrell, 1995] and has since become a main development platform for interactive experiences.



Figures 1 and 2. The IVE space

### *Scripted Applications*

My first project in the IVE space was an interactive story/museum-exhibit called Encounters. A member of the public would meet a 3D humanoid character at a crossroad of a 3D virtual museum-city. S/he would be handed a message and become involved in solving a mystery regarding three contemporary artists. Solving the mystery, brought the participant through a series of chambers, and made him/her become familiar with the work of the artists. The person would interact with the characters, sounds, and images

projected on a large screen through simple gestural and voice commands. This project was entirely scripted and – although I managed to author some interesting segments – its authoring complexity soon grew to a size which was very hard to handle for the designers or the participant and nonetheless it was too simple for the public to be able to enjoy and appreciate.

By attempting to author a real time sensor-driven application as a scripted system I learned:

- the interaction with the public is too rigid: it obliges participants to perform a pre-determined action in order to move on to the follow-on presentation segment.
- it is hard to reconfigure: whenever I needed to change the plot I had to re-author the whole system.
- the authoring complexity of the system grew exponentially, and as a consequence the final interactive narrative ended up being so simple that its shallow depth of content soon became uninteresting to the public.

### *Responsive Applications*

In February 1996 I shifted my attention from storytelling to performance and, having learned from my previous experience, I aimed for a system which would be uncomplicated to understand and use. I created *DanceSpace*: an interactive stage for a single performer [figures 3,4,5] in which music and graphics are generated on the fly by the dancer's movements. In *DanceSpace* [Sparacino, 2000b] a small set of musical instruments is virtually attached to the dancer's body and generates a melodic soundtrack in tonal accordance with a soft background musical piece. The performer projects graphics on a large back screen using the body as a paintbrush. In *DanceSpace* both common users and performers are usually able to quickly understand the interface and choreograph improvisational pieces influenced by the technological opportunity. *DanceSpace* is a typical example of a responsive experience with one-to-one mappings between sensory input – the dancer's hands/feet/head/center of body movements – and system output –music and graphics.



Figure 3. Performers in DanceSpace



Figures 4 and 5. Body tracking system used for DanceSpace and the body driven City of News

I also authored City of News with a responsive architecture. City of News is a dynamically growing urban landscape of information. It is an immersive, interactive, web browser that takes advantage of people's strength remembering the surrounding three-dimensional spatial layout. Starting from a chosen "home page", where home is finally associated with a physical space, our browser fetches and displays URLs so as to form skyscrapers and alleys of text and images through which the user can navigate. In the responsive version of City of News the browser is driven by body gestures and by the position of the user on the projected floor map, both identified by a computer vision system running in real time. The system recognizes the following commands "*scroll up*" both arms up, "*scroll down*" both arms down, "*next*" right arm stretched, "*previous*" left arm stretched, and "*follow link*" given by the position of the body on the map. The system uses these body gestures to trigger browsing actions with the one-to-one mapping described.



Figures 6 and 7. Visitors using the body driven City of News at SIGGRAPH 99



With DanceSpace and City of News I learned:

- Authoring interactive applications with one to one mapping between user input and system output is much easier than authoring a scripted system: the authoring complexity is reduced to generating a look up table which associates inputs with outputs
- Using these systems can be very rewarding for the public, if the input-output mapping is simple and easy enough that most people can guess it and apply it within the first exploratory minutes with the system.
- While responsive systems are easy to author, they are nevertheless prone to error: users perform gestures in different ways, and lacking a history of the interaction or other contextual information, the system is prone to misclassify human gestures and to produce the wrong response, no matter how good the gesture classification system may be.
- There is no perfect sensor, nor a perfect gesture classifier, therefore the noise intrinsic to the sensor measurements as well as the imperfection of the classifier, is also a source of unreliability of the system. The result is that the public becomes confused about which gestures causes which response, and soon becomes disengaged from the interactive experience
- While authoring is easy, the authoring complexity is still limited. With this authoring is it hard to articulate a narration which has some depth or development, such as one which has a simple introduction-development-conclusion, as the geography of the input-output mapping imposes only a stimulus-response coupling of human gestures with system responses.

### *Behavioral Applications*

Virtual Studio: Digital Circus was first constructed in March 1997. It is an immersive behavioral experience in which all objects present in the 3d virtual circus are endowed with behaviors. Advanced real-time computer vision techniques allow the system to composite and blend a 2d image of the participant inside the 3d world, without the need of blue screens. Individual distant participants can be remotely connected to and share the same virtual world [figures 8,9,10]. Hence such setup can be used at home for collaborative storytelling, visual communication from remote locations, or game playing. In the circus a behavior-based butterfly pet follows the participant around, the cannon fires a cannon woman when the participant virtually presses a virtual button, an umbrella appears at need. Sitting on a chair causes a gramophone to appear and music to be played. An arm gesture causes the participant to grow taller or become tiny-small, on request. All of these actions/transformations are possible because each object in the virtual space is endowed with an autonomous behavior and it takes care of doing the right thing at the right time.





Figures 8,9 and 10. Visitors using the participative City of News which extends Virtual Studio/Digital Circus to the 3D web as the virtual set

With Virtual Studio/Digital Circus I learned that:

- Behavioral applications overcome many of the limitations of both scripted and responsive systems.
- They are limited as they do not adapt nor learn from the user: they are therefore unable to tailor content to specific participants or types of participants
- As with responsive systems, they are unable to understand the user's gestures in context and are prone to error in interpreting human actions.

By grouping some of my previous work according to the taxonomy described above and by analyzing the interaction modality and experience of the public in each application, I derived the following conclusions:

- Scripted experiences are difficult to author. They require a careful and detailed planning of the material presented. Complexity burdens both the author and the recipient of the piece. If the interactive experience requires this approach - such as in the case of some storytelling projects - it is important to keep the project small and simple.
- Responsive experiences can be successful, especially when the system responds in a timely fashion. Also it is important that the input-output mapping can be made clear to the public as early as possible in the course of the experience (City of News) for the public to be engaged rather than confused by the interactive application. However, due to the fact that the input-output mapping is invariant,

responsive applications can become repetitive and obsolete after a few experiences.

- Behavioral experiences allow the public to experience more complex forms of interaction. The behavior architecture allows distribution of the authoring complexity to the various characters/media in the piece. Once the behavior system is constructed, such systems are much easier to build than the scripted pieces, as they require only specifications about the behavior parameters for the specific application considered.

However, if we want to build experiences which can articulate a digital audiovisual narration as a function of the user's body movements or path, as perceived by the sensors, we need alternative architectures which can understand the visitor's intentions in context, and are robust with respect to the variability of user input and the noise intrinsic in the sensors' measurements.

## **2.3. Requirements for new authoring tools**

Developing the previously described projects I made a series of observations which have become the motivation and ground for the research described in this document. Based on the previous analysis, I summarize in this section the main requirements that new authoring tools need to have to be used effectively in real time interactive spaces.

### ***Robust sensing***

Robust sensing is the premise for the correct interpretation of the user's intention. Without it, we would have a system which, like some old grandmothers, is slightly deaf, and produces answers which sometimes do not make sense with the question asked. This happens not because the system/grandmother is not smart and knowledgeable and interesting, but because the sensorial percept (question asked) was modified and perturbed by an imperfect sensor (the ear) and therefore the system/grandmother misinterpreted the input and produced an answer to a different question than the one asked. Mono-sensor applications which rely on one unique sensor modality to acquire information about the participant are brittle and prone to error. For how well that one sensor works individually, whether that be a camera, or a radar, or an electric field sensor, it only provides the system with a single view of what is going on. In order for a body driven interactive application to offer reliable and robust response to a large number of people on a daily basis in a museum, or meet the challenges of the variable and unpredictable factors of a real life situation, we need to rely on a variety of sensors which cooperate to gather correct and reliable measurements on and about the user. Cooperation of sensor modalities which have various degrees of redundancy and complementarity can guarantee robust, accurate perception. We can use the redundancy of the sensors to register the data they provide with one another. We then use the complementarity of the sensors to resolve ambiguity or reduce error when an environmental perturbation affects the system.

### ***Interpretation of data***

To make good use of reliable measurements about the user, we need to be able to interpret our measurements in the context of what the user is trying to do with the digital media, or what we actually want people to do to get the most out of the experiences we wish to offer. The same or similar gesture of the public can have different meanings according to the context and history of interaction. For example the same pointing gesture of the hand can be interpreted either as pushing a virtual character, or more simply, as a selection gesture. In a similar way, the system needs to develop expectations on the likelihood of the user's responses based on the specific content shown. These expectations influence in turn the interpretation of sensory data. Following on the previous example, rather than teaching both the user and the system to perform or recognize two slightly different gestures, one for pushing and one for selecting, we can simply teach the system how to correctly interpret slightly similar gestures, based on the context and history of interaction, by developing expectations on the probability of the follow-on gesture. In summary, our systems need to have a user model which characterizes the behavior and the likelihood of responses of the public. This model also need to be flexible and should be adaptively revised by learning the user's interaction profile.

### ***Compelling response***

It is difficult to produce compelling applications simply by direct mapping of sensor measurement inputs with digital media output. While this strategy may work for very simple interactive environments, it is not effective for producing an engaging application. I would like to use the term compelling multimedia application in analogy to what in computer graphics researchers call a believable synthetic character. A believable character is one whose response to the user's input is appropriate to its role and history of interaction. Appropriate responses are those that make sense, or that the user can make sense of, such as for an interactive dog fetching the ball when the user throws it, or sit and rest when sleepy. By analogy in multimedia we want applications that are convincing and not repetitive or shallow. To describe this feature I use the term compelling response. Many current interactive systems are defined by a series of couplings between user input and system responses. The problem with these systems is that they are often repetitive: the same action of the participant always produces the same response by the system. Alternatively, most existing CDROM titles are scripted: they sequence micro-stories in multi-path narrative threads. Examples of such titles are the popular CD-ROM based game MYST [<http://sirrus.cyan.com/Online/Myst/MystHome>] or Disney's Hunchback of Notre Dame Storybook CD-ROM. While the content presentation in these applications tends to be more engaging, they often impose a rigid interaction modality and become boring after a while. The participant's role is confined to clicking and choosing the sequencing of the narrative thread without real engagement or participation in the narrative. In order to create compelling interactive environments we need to be able to simulate encounters between the public and the digital media acting as a character. To accomplish this goal we need to be able to model the story we wish to narrate in such a way that it takes into account and

encompasses the user's intentions and the context of interaction. Consequently the story should develop on the basis of the system's constant evaluation of how the user's actions matches the system's expectations about those actions, and the system's goals.

Sto(ry)chastics, described in the following chapters, offers a powerful tool to model a real time sensor driven interactive storyteller according to the parameters mentioned above. When modeling user-story interaction we need to preserve on the one hand the causal effects of user over story or of story on the user. At the same time, we need to be able to account for the variability in interaction, i.e. the improvisational nature of interaction I described in section 1.2. As I will show in the following chapters of this document, sto(ry)chastics allows the interactive experience designer to have flexible story models, decomposed in atomic or elementary units, which can be recombined into meaningful sequences at need in the course of interaction. Another reason to use sto(ry)chastics to understand the user's intention is that it allows us to model both the noise intrinsic in interpreting one's intentions in general as well as the noise intrinsic in telling a story. We as humans do not tell the same story in the same way all the time, and we naturally tend to adapt and modify our stories to the age/interest/role of the listener. The Bayesian network approach is therefore the most apt to model noisy sensors, noisy interpretation of intention, and noisy stories.

## 2.4. Related Work

Oliver [http://www.media.mit.edu/~nuria/dypers/dypers.html; Schiele, 1999] developed a wearable computer with a visual input as a visual memory aid for a variety of tasks, including medical, training, or education. This system records small chunks of video of a curator describing a work of art, and associates them with triggering objects. When the objects are seen again at a later moment, the video is played back. The museum wearable differs from the previous application in many ways. DYPERS is a personal annotation device, and as opposed to the museum wearable, it does not attempt to perform either user modeling or a more sophisticated form of content selection and authoring. It does one-to-one associations between triggering objects and recording or playout of clips. Besides general training, is used specifically in the museum context to allow a visitor to record salient moments of the explanation by a human guide to later replay them in the context of an independent visit to a museum, without a guide. The museum wearable in contrast focuses on estimating the visitor's type and interest profile to deliver a flexible user-tailored narrative experience from audio/video clips that have been prerecorded. These clips or animations would usually be part of the museum's digital media collection. As opposed to DYPERS, it does not have the ability to record new content for it to be played out at a later time. Its purpose is to create for the visitor a path-driven personalized and immersive cinematic experience, which takes into account the overall trajectory of the visitor in the museum, the amount of time that visitors station to look at, and explore the objects on display, to select a personalized story for the visitor, out of several possible digital stories that can be narrated.

Feiner [Feiner, 1997] has built a university campus information system, worn as a wearable computer. This device is endowed with a variety of sensors for head tracking and image registration. Both the size of the wearable, mounted on a large and heavy backpack, as well as the size of the display, are inappropriate for it to be used for a museum visit.

The author showed an early prototype of the museum wearable, based on the above architecture, as a demonstration for the SIGGRAPH 99 Millennium Motel [figures 6,7], and received outstanding feedback and encouraging comments from the audience [Sparacino, 1999; Sparacino, 2000a].

In addition to the specific research on wearables, to carry out the research described in this document, I have used knowledge from various disciplines: interactive computer graphics, statistical modeling and Artificial Intelligence (probabilistic reasoning, Bayesian networks), user modeling, wearable computing, and probabilistic knowledge representation for content organization and delivery. The task of interpreting the visitor's intentions and desires during the museum visit is similar to traffic surveillance: understanding driver behavior from external sensors (cameras, radars) placed along the major streets and intersection, a field of research which has recently successfully been addressed using probabilistic Bayesian networks (what is of interest in traffic surveillance research, is the early identification of aggressive driver types before they cause major street accidents). I describe in this section some of the work in the above mentioned fields that has partially guided and inspired my research on sto(ry)chastics and the design of the museum wearable. I give some weight on the interactive computer graphics background section, as this is the field in which I have mostly developed my previous work.

### ***Interactive Computer Graphics***

Blumberg and Galyean [Blumberg and Galyean, 1995] use an ethological model to build behavior-based graphical creatures capable of autonomous action, and who can arbitrate response to external control and autonomy. They introduce the term “directability” to describe this quality. Hayes-Roth [Hayes-Roth, 1996] uses the notion of directed improvisation to achieve a compromise between “directability” and life-like qualities. Her research aims at building individual characters that can take directions from the user or the environment, and act according to these directions in ways that are consistent with their unique emotions, moods, and personalities (improvisation). Magnenat Thalmann and Thalmann [Magnenat Thalmann and Thalmann, 1993] have built a variety of examples of virtual humans equipped with virtual visual, tactile, and auditory sensors to interact with other virtual or real (suited/tethered) humans [Emering, 1997]. In [Perlin, 1996] Perlin describes an authoring system for movement and action of graphical characters. The system consists of a behavior engine which uses a simple scripting language to control how actors communicate and make decisions, and an animation engine which translates programmed canonical motions into natural noisy movement. Terzopoulos provided a fascinating example of behavior based graphical fishes endowed with synthetic vision and which can learn complex motor skills [Terzopoulos 1994, 1999]. Tosa has built characters which can understand and respond to human emotion using a combination

of speech and gesture recognition [Tosa, 1996]. Bates and the Oz group have modeled a “woggles” world inhabited by woggles with internal needs and emotions and capable of complex interactions with the user [Bates, 1992]. Their user’s interface is though mainly limited to mouse and keyboard input.

### ***Bayesian Networks***

The work of Pearl [Pearl, 1988] is fundamental to the field of Bayesian networks. Jordan’s book [Jordan, 1999] had the merit of grouping together some of the major advancements since Pearl’s 1988 book. Cowell et al. [Cowell, 1999] provide a comprehensive up-to-date introduction to Bayesian networks. Jensen [1996; 2001] has written two thorough introductory books that provide a very good tutorial, or first reading, in the field. Bayesian networks have gained popularity in the early nineties, when they were successfully applied to medical diagnosis [Heckerman, 1990]. More specific references to Bayesian networks will be given in Chapter 4, and are embedded appropriately in the text as I present and describe Bayesian networks.

### ***User Modeling for Computer Games***

Albrecht et al [Albrecht, 1997], have been amongst the first to model the behavior of a participant to a computer game using Bayesian networks. Jebara [Jebara, 1998], uses CHMMS, which are a particular case of Dynamic Bayesian Network, to perform, first analysis, and then synthesis, of a player’s behavior in a game. Brand, Oliver, and Pentland [Brand, 1996; Brand, 1997], also use a Coupled Hidden Markov Models approach to successfully recognize Tai-Chi gestures in the context of a Tai-Chi training game.

### ***Learner Modeling***

[Henze, 1999] uses Bayesian networks to assess the stating knowledge of a learner and builds an adaptive hypermedia system using a constructivist approach. Conati et al. [Conati, 1997] have built an intelligent tutoring system able to perform knowledge assessment, plan recognition and prediction of students’ actions during problem solving using Bayesian networks. Jameson [Jameson, 1996], provides a useful overview of student modeling techniques, and compares the Bayesian network approach with other popular modeling techniques such as fuzzy logic.

### ***Traffic Surveillance***

Forbes et al. [Forbes, 1995] use an agent based belief network and agent centered features to recognize driving activity from simulated and real data. Kwon and Murphy [Kwon and Murphy, 2000], use coupled hidden markov models to learn a model of traffic velocities from data for fault diagnosis and to predict future traffic patterns. Pynadath and Wellman [Pynadath and Wellman, 1995], use a Bayesian network approach to induce the plan of a driver from observation of vehicle movements. Starting from a model of how the driver generates plans, they use highway information as context that allows the system to correctly interpret the driver’s behavior.

### ***Wearable Computing***

Starner et al [Starner, 1997], describe seminal work in wearable computing. Behringer et al. [Behringer, 1999] group a variety of augmented reality techniques spanning from real time computer vision registration to industrial and medical applications. In: Sensing Techniques for Mobile Interaction, Hinckley et al [Hinckley, 2000] discuss tradeoffs between real time sensing and traditional user interface approaches to ease execution of common tasks on handheld devices.

### ***Probabilistic Knowledge Representation***

Koller and Pfeffer [Koller and Pfeffer, 1998] have done innovative work in using probabilistic inference techniques that allows most of the frame bases knowledge representation systems available today to annotate their knowledge bases with probabilistic information, and to use that information to answer probabilistic queries. Their work is relevant to describe and organize content in any database system so that it can later be selected either by a typed probabilistic query or by a sensor driven query. Pasula and Russell [Pasula and Russell, 2001] describe efficient Markov Chain Monte Carlo techniques to handle reference uncertainty and identity uncertainty for relational probability models. Pfeffer, Koller, and al. [Pfeffer, 1999] have also provided an example of application of probabilistic object-oriented knowledge representation in: SPOOK: a system for military situation assessment for battleships.

## **Chapter 3**

# **Application: the Museum Wearable**

### **3.1. Scenario**

When walking through a museum there is so many stories we could be told. Some of these are biographical about the author of an artwork, some are historical and allow us to comprehend the style or origin of the work, and some are specific about the artwork itself, in relationship with other artistic movements. Museums usually have large web sites with multiple links to text, photographs, and movie clips to describe their exhibits. Yet it would take hours for a visitor to explore all the information in a kiosk, to view the VHS cassette tape associated to the exhibit and read the accompanying catalogue. Most people do not have the time or motivation to devote to assimilate this type of information, and therefore the visit to a museum is often remembered as a collage of first impressions produced by the prominent feature of the exhibits, and the learning opportunity is missed. How can we tailor content to the visitor in a museum, during his/her visit, to enrich both the learning and entertaining experience ? We want a system which can be personalized to be able to dynamically create and update paths through a large database of content – such as the one already developed for interactive kiosks – and deliver to the user during the visit all the information he/she desires in real time. If the visitor spends a lot of time looking at a Monet, the system needs to infer that the user likes Monet and will update the paths though the content to take that into account. By doing so it will update the user's profile and the path through the content to be delivered. This thesis proposes stochastic story modeling as an effective way to turn this scenario into reality.

### **3.2. The Museum Wearable**

In the last decade museums have been drawn into the orbit of the leisure industry and compete with other popular entertainment venues, such as cinemas or the theater, to attract families, tourists, children, students, specialists, or passerbiers in search of alternative and instructive entertaining experiences. Some people may go to the museum for mere curiosity, whereas other may be driven by the desire of a cultural experience. The museum visit can be an occasion for a social outing, or become an opportunity to meet new friends. While it is not possible to design an exhibit for all possible categories of visitors, it is desirable to offer the exhibit designers methods by which it becomes possible to augment the visitors' knowledge about the exhibit in a more personalized way for the different people or visitor categories. This thesis focuses on the museum wearable, and a more general purpose authoring technique, called sto(ry)chastics, as a technological



aid which help the curator better communicate with the museum audience, and the visitor to possibly have a more engaging, instructive, and entertaining experience. The research described in this document illustrates the hardware, authoring techniques, and software created for the construction of the museum wearable, and points to other possible applications of the same authoring techniques. It postpones, as future work, the assessment at the exhibit's site, of the museum wearable's contribution to the public's experience and its ability to facilitate new exhibit design.

Traditional storytelling aid for museums have been: *signs and text labels*, spread across the exhibit space; exhibit *catalogues*, typically sold at the museum store; *guided tours*, offered to groups or individuals; *audio tours*, and more recently video or multimedia *kiosks* with background information on the displayed objects. Each of these storytelling aids have advantages and disadvantages. Catalogues are usually attractive and well done, yet they often too are cumbersome to carry around during the visit as a means to offer guidance and explanations. Guided tours take away from visitors the choice of what they wish to see and for how long. They can be highly disruptive for the surrounding visitors, and their effectiveness strictly depends on the knowledge, competence, and communicative skills of the guide. Audio tours are a first step to help augment the visitor's knowledge. Yet when they are button activated, as opposed to having a location identification system, they can be distracting for the visitor. The information conveyed is also limited by the medium: it's only audio. It is not possible to compare the artwork described with previous relevant production of the author, nor show other relevant images. Interactive kiosk are more frequently found today in museum galleries. Yet they are physically distant from the work they describe thus not supporting the opportunity for the visitor to see, compare, and verify the information received against the actual object. My personal experience suggests that when extensive web sites are made available through interactive kiosks placed along the museum galleries, these may absorb lengthy amount of time from the visitor's museum time, thereby detracting from, rather than attracting to, the objects on display. Finally panels and labels with text placed along the visitors' path can interrupt the pace of the experience as they require a shift of attention from observing and contemplating to reading and understanding [Klein, 1986].

Another challenge for museums is that of selecting the right subset of representative objects among the many belonging to the collections available. Usually, a large portion of interesting and relevant material never sees the light because of the physical limitations of the available display surfaces.

Some science museums have been successfully entertaining their public, mainly facilitated by the nature of the objects they show. They engage the visitor by transforming him/her from a passive viewer into a participant by use of interactive devices. They achieve their intent, amongst other things, by installing button-activated demonstrations and touch-sensitive display panels which provide supplementary information when requested. They make use of proximity sensors to increase light levels on an object when a visitor is close-by and/or to activate a process. Other museums – especially those which have large collections of artwork, like paintings, sculptures, and manufactured objects -- use audiovisual material to give viewers some background and a coherent narrative of the works they are about to see or that they have just seen. In some cases, they provide audio-tours with headphones along the exhibit. In others, they

dedicate sections of the exhibit to the projection of short audiovisual documentaries about the displayed material. Often, these movies that show artwork together with a description of their creation and other historical material about the author and his times, are even more compelling than the exhibition itself. The reason is that the documentary has a well edited story, and the visuals are nicely orchestrated and come with music and dialogues. The viewer is then offered a more unified and coherent narration than in the fragmented experience of the visit. A visit to a museum demands, as a matter of fact, a certain amount of effort, knowledge, concentration, and guidance, for the public to leave with a consistent and connected view of the material presented.

Technology can help the visitor reconstruct a connected story for the objects on display at the museum, and offer curators way of addressing different visitor categories, by creating experiences in which the objects themselves narrate their own story in context. Wearable computers have recently raised to the attention of technological and scientific investigation [Starner, 1997] and offer an opportunity to “augment” the visitor and his perception/memory/experience of the exhibit in a personalized way. This thesis presents a wearable computer which dynamically edits a documentary about the shown objects according to the path of the visitor inside the physical space of the museum. The museum wearable [figure 11] targets individual visitors with special learning needs or curiosity, and offers a new type of entertaining and informative museum experience, more similar to immersive cinema than to the traditional museum experience. It selects and presents audiovisual sequences from a documentary database and adapts them to the visitor’s type and profile it estimates during the visit. The user modeling process is obtained with a Bayesian network using as input the information provided by the location identification sensors on where and how long the visitor stops. The content selection process is also influenced by the constraints on segment ordering provided by the curator to ensure the assembly of a “good” story.

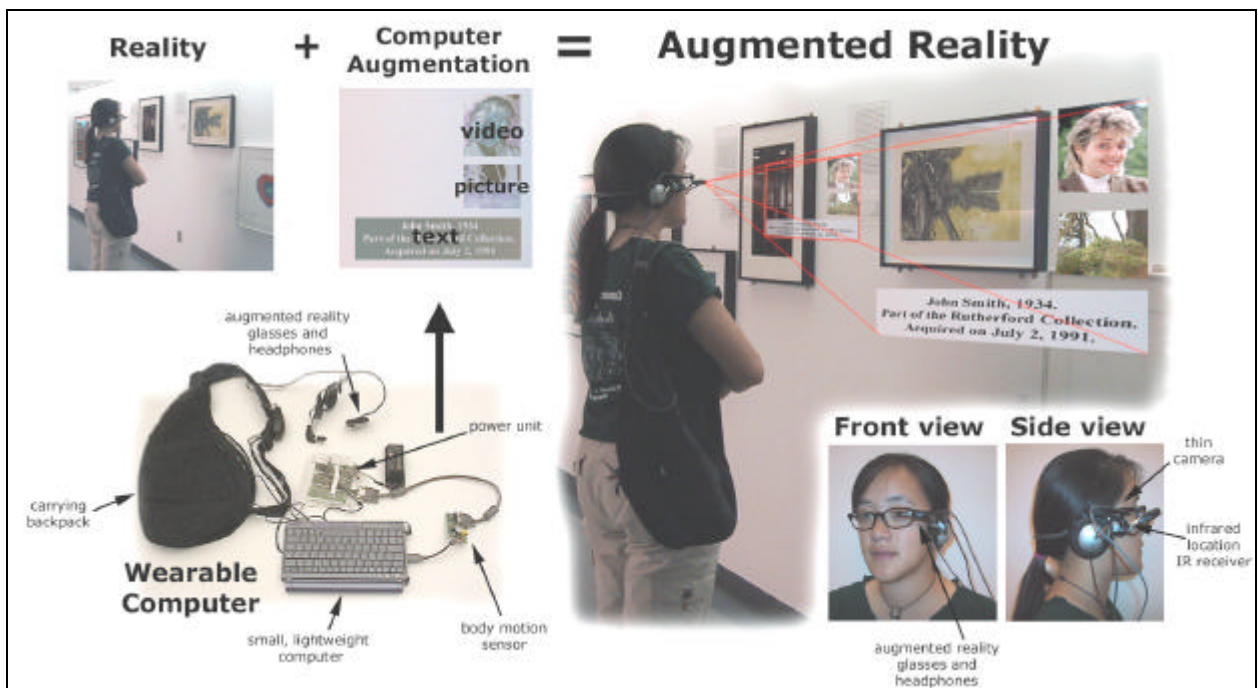


Figure 11. The Museum Wearable: explanation of concept and application

### 3.3. Visitor Types

Museums are designed to offer people a curated experience. An understanding of audience needs and expectations is a primary mission of today's museums. This leads to identifying first of all groups of individuals who share common traits such as culture, leisure preferences, fields of studies, ethnic or social affiliations, disabilities, socio-economic levels, and so forth. Any identifiable sub-group within a community is potentially a museum target audience. Secondly the exhibit curator and designer need to assess the basic knowledge and expectations of any of these subgroups to be able to reach, communicate, and stimulate curiosity in their visitors. Dean [Dean, 1994] cites the Arnold's Values and Lifestyles Segments (VALS) model as a useful tool to identify target audiences [ibidem, p. 21]. This model classifies people along a psychological maturity scale which includes need driven (survivors, sustainers), Outer driven (belongers, emulators, achievers, integrated) and inner directed (experiential, I-am-me, societally conscious) individuals.

Eleanor Hooper-Greenhil identifies target groups which include: families, school parties, other organized educational groups, leisure learners, tourists, the elderly, and people with visual, auditory, mobility or learning disabilities [Hooper-Greenhil, 1999, p. 86]. She then suggests a partition of museum resources, to target, attract and entertain these different groups. This is of interest to our research, as it provides grounding for partitioning the largely available digital content resources for the same exhibit, according to the visitor types that we have chosen to identify with the museum wearable, as described in the next chapter.

During a personal interview, Beryl Rosenthal, director of exhibitions at the MIT Museum, described a more sophisticated visitor type classification. She identified: stroller moms, accompanied by children three years old or younger, window shoppers: families who cruise through the museum in search of an alternative leisure experience, button pushers (when buttons available): typically adolescents, school groups, the date crowd, the phds, who want to know (and criticize) everything in the museum. Young visitors, children in the 5-14 also represent a separate group of visitors with different learning needs and curiosities than the other groups. While this colorful classification well depicts the variety of public that museums need to equally attract, entice and educate, it is too sophisticated to model mathematically, at least initially.

More usefully for this research, Dean [Dean, 1994] generalizes museum visitors in three broad and much simpler categories [ibidem, pp. 25-26]. The first category includes what he calls the "casual visitors": people who move through a gallery quickly and who do not become heavily involved in what they see. Casual visitors use some of their leisure time in museums but do have a strong stimulus or motivation to deepen their knowledge about the objects on display. The second group, the "cursory visitors" show instead a more genuine interest in the museum experience and their collections. According to Dean these visitors respond strongly to specific objects that stimulate their curiosity and wander through the gallery in search of further such stimulus for a closer exploration of the targeted objects. They do not read every label nor absorb all available information, but will occasionally read and spend time in selected areas or with selected objects of interest they encounter in the galleries. The third group is a minority of visitors who

thoroughly examine exhibitions with much more detail and attention. They are learners who will spend an abundance of time in galleries, read the text and labels, and closely examine the objects.

Dean attributes differences between “people who rush”, “people who stroll”, and “people who study” to different prior experiences and educational level. Yet he states that it is important that museums are equipped to communicate and interest all visitors, by scaling and designing an exhibit so that it offers entertainment to the “stroller” as well as an opportunity to deepen knowledge for the “learners”.

Serrell, B. [Serrell, 1996] also divides visitors into three types: the transient, the sampler and the methodical approach to viewing. She notes that currently museum evaluators are using terms like “streakers, studiers, browsers, grazers and discoverers” to characterize museum visitors’ styles of looking at exhibits. But she concludes that this type of categorization is not useful for summative evaluation, suggesting that it is a subjective method of classification, and that it is not fruitful to try and create exhibitions that serve these different styles of visiting. Instead she suggests that a more objective means of classification be found, such as average time spent in the exhibition space. Her studies suggest that visitors cover 500 square feet per minute in exhibition spaces, a figure she finds somewhat frustrating—since it is too fast for visitors to actually be reading and absorbing the material. Also disappointing is that the average visitor uses about 30% of the exhibit. Given that visitors are going to view only a small percentage of the exhibit, Serrell suggests that developers build redundancy into the framework of the show, simplifying concepts and repeating them in different ways throughout the space. Developers should ask themselves: Does this exhibit make sense and is it likely to be meaningful for any self-selecting visitor who stops at it? They should also think about a one-time, time-limited non sequential motivated but non expert person as their audience.

For more details on results, methods and approaches for visitor studies, the author forwards the reader to the US’s Visitor Studies Association (VSA), which provides a forum for the exchange of information in the field of visitor studies [<http://museum.cl.msu.edu/vsa/whoware.htm>]. The VSA conducts three types of research: interviewing visitors, tracking (time spent in different areas of the exhibit), and observing (scans or sweeps). Founded in 1992 it provides information on development of methodology for visitor studies, visitor surveys and audience development and though its annual conference is one of the major sources for understanding visitor behavior in the changing mission and landscape of the contemporary museum.

In accordance with the simplified museum visitor typology suggested by Dean and Serrell, I have chosen to identify three main visitor types using a Bayesian network. To offer a more intuitive understanding of the types described by Dean and Serrell I have renamed them: the busy, selective, and greedy visitor type. Should it be necessary or desirable, the identification of other visitor types or subtypes has been postponed to future improvements and developments of this research.

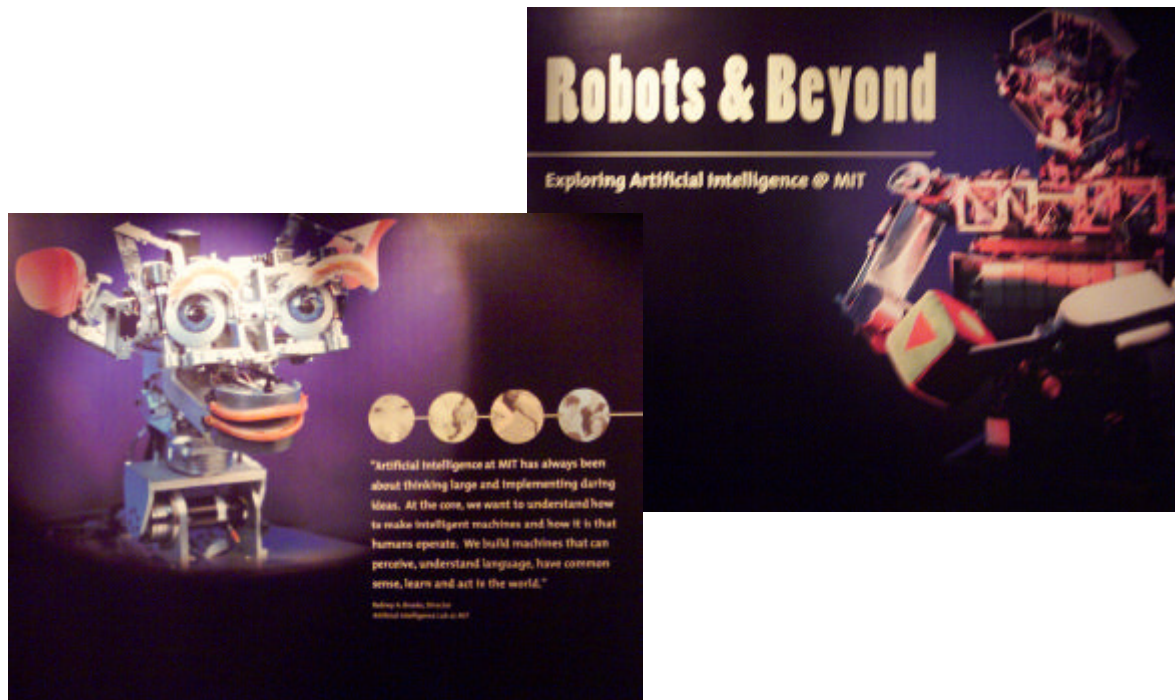
### 3.4. Experimentation Platform: the Museum Wearable at MIT Museum's "Robots and Beyond" exhibit

The ongoing robotics exhibit at the MIT Museum provided an excellent platform for experimentation and testing with the museum wearable. This exhibit, called Robots and Beyond, features landmarks of MIT's contribution to the field of robotics and Artificial Intelligence. The exhibit is organized in five sections: Introduction, Sensing, Moving, Socializing, and Reasoning and Learning, each including a few robots, a video station, and posters with text and photographs which narrate the history of robotics at MIT [figures 13,14,15,16,17]. There is also a large general purpose video station with large benches for people to have a seated stop and watch a PBS documentary featuring robotics research from various academic institutions in the country.

The posters and labels occupy half of the available exhibit wall space, and while they certainly provide useful information, they require long stops for reading, take useful space away from other interesting objects which could be displayed in their stead, and are not nearly as compelling and entertaining as a human narrator (a museum guide) or a video documentary about the displayed artwork. On the other hand, the video stations, located in each section of the exhibit, complete the narration about the artwork by showing the robots in motion and by featuring interviews with their creators. While the video stations provide compelling narrative segments, they are not always located next to the object described, and therefore the visitor needs to spend some time locating the described objects in the surrounding space in order to associate the object to the corresponding narrative segment. The video stations detract attention from the actual objects on display, and are so much the center of attention for the exhibit that the displayed objects seem to be more of a decoration around the video stations than being the actual exhibit.



Figure 12. Visitor testing the Museum Wearable at MIT Museum's *Robots and Beyond* Exhibit



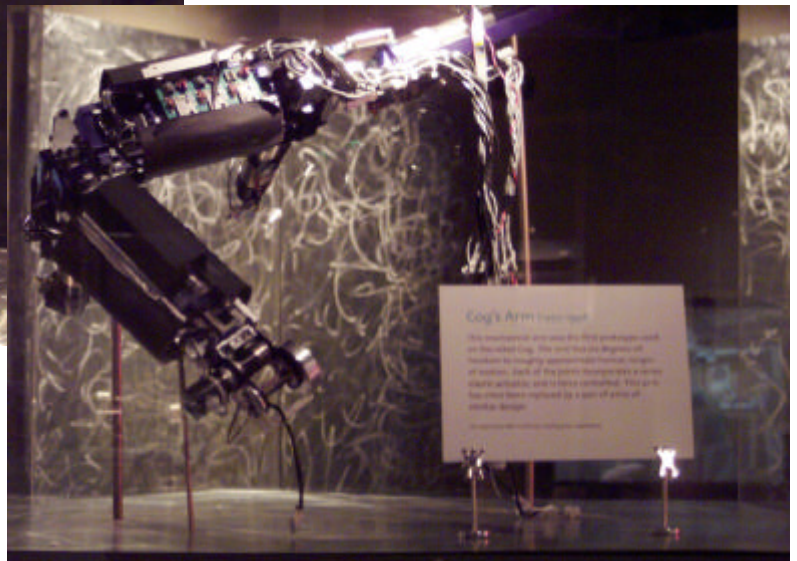
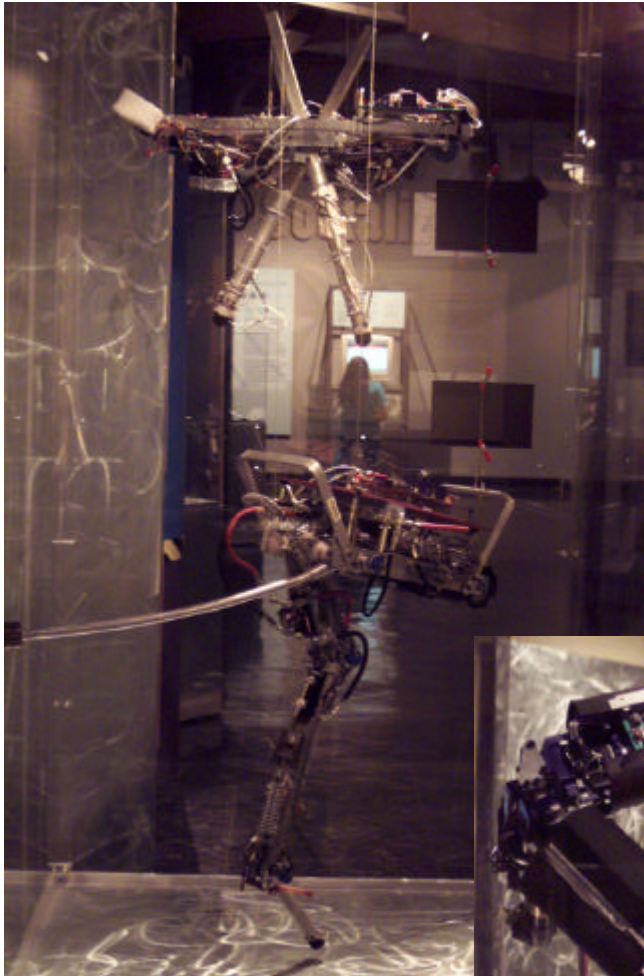
Figures 13 and 14. Images from MIT Museum's *Robots and Beyond* Exhibit

By providing an audiovisual narration synchronized to the objects on display, and tailored to the visitor's interests, the museum wearable proposes a new device which will hopefully transform and enhance the visitor's experience. Some of the information and narration originally provided by the posters and video stations is instead provided by the wearable, and it is edited and choreographed along the visitor's path, virtually placed next to the objects it describes [figure 12]. The first immediate advantage of designing the exhibit to be viewed with the wearable is a much larger availability of space to display a larger and more complete variety of objects. The author's expectation is that with the wearable the visitor's attention can be focused directly onto the objects, and their path will not be interrupted by long stops reading posters and watching the TV monitors placed along the museum space. While further testing of the museum wearable's effect on the public will have to be carried out as future work, similar considerations could be made for interactive kiosks which would be no longer needed allowing more room for additional objects. Currently, at MIT's *Robots and Beyond* exhibit, the kiosks interrupt the flow and pace of the visitor's path, and the information they offer could be more effectively delivered in smaller chunks along the visitor's path inside the museum, and next to the objects they describe.

To better tailor the museum wearable to the public, and to preserve and possibly enhance the original message of the exhibit, I interviewed Janis Sacco, one of the main curators for *Robots and Beyond*. The curator said that one of the main goals of the exhibit is to stimulate the public to think about what intelligence is and to show how scientists can successfully emulate some aspects of human intelligence. This explains why the exhibit is organized in sections that highlight the different facets of human intelligence, such as perceptual intelligence (sensing), motor and emotional intelligence, and, finally,



high level reasoning. Janis said that she expected visitors to leave with an understanding the complexity of living organisms in general and she also noted that it is quite hard to capture some of that complexity into machines capable of imitating simple human or animal behavior. When asked if she would welcome the museum wearable as an aid and guide to better understand and appreciate the exhibit she enthusiastically asserted: “It is the interactivity that we lack. What you are doing is testing a device that is made with AI in an exhibit that is about AI and using it to let people experience AI.” I read this as an encouragement to go on with the research described here below.



Figures 15,16 and 17. Images from MIT Museum's *Robots and Beyond* Exhibit

### **3.5. Annotations and observations of visitors' behavior**

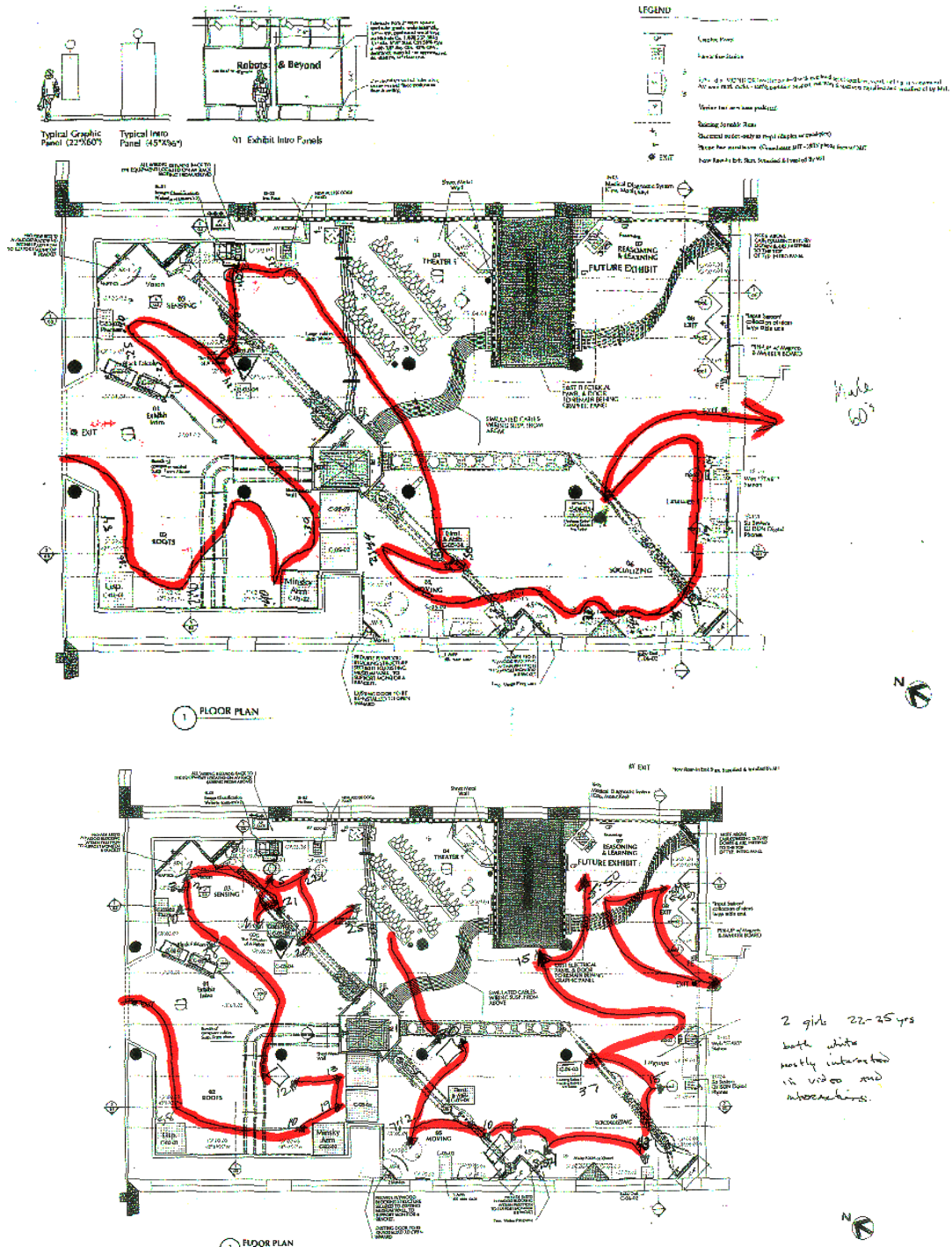
In order to have the museum wearable understand the visitors' interests, I first gathered experimental data on the visitors' behavior in the museum. This was an important preliminary step, as the starting point in building a Bayesian network is usually to model the expert knowledge about the domain, and to assign prior probabilities to the nodes of the network on the basis of this knowledge. I discuss in the next chapter the steps undertaken to build the mathematical model for the visitor. In this paragraph I present the methodology and results that led us to have a quantitative assessment of the visitors' behavior at the MIT museum's Robots and Beyond exhibit. According to the VSA (cfr. 4.2.), timing and tracking observations of visitors are often used to provide an objective and quantitative account of how visitors behave and react to exhibition components. This type of observational data suggests the range of visitor behaviors occurring in an exhibition, and indicates which components attract, as well as hold, visitors' attention. Usually in the case of a complete exhibit evaluation this data is accompanied by interviews with visitors, before and after the visit.

For Robots and Beyond, the curator Janis Sacco, shared her findings resulting from interviews with the visitors, and I could therefore focus uniquely in gathering tracking data. I followed the classic approach: during the course of several days I had a team of people at the MIT Museum individually track and make annotations about the visitors. Each member of the tracking team had a map and a stop watch. Their task was to draw on the map the path of individual visitors, and annotate the locations at which visitors stopped, the object they were observing, and how long they would stop for. In addition to the tracking information, the team of evaluators was asked to assign a label to the overall behavior of the visitor, according to the three visitor categories identifies by Dean, that I described earlier, and which I renamed "busy", "greedy", and "selective". Together with the curator, and in accordance to the literature, I have found that allowing the evaluators to make a subjective judgment about the visitor's behavior is as accurate as asking the visitors themselves. Visitors who are not familiar with the description of the three categories described by Dean would tend to misclassify themselves. In addition to that the museum wearable acts as an external observer, who tailors a personalized story to the visitor, on the basis of external observations, as opposed to asking the visitor what they want to do at every step. Lastly, the assessment made by the team of evaluators is used to initialize the Bayesian network, but the model can later be refined, that is the parameters can be fine tuned, as more visitors experience the exhibit with the museum wearable, as described in the next Chapter.

The visitor tracking information is shown in table 1. I tracked about 50 visitors, and gathered 50 such tracking sheets from the team of evaluators [figures 18,19]. The data they tracked is summarized in table 1. The table contains raw data, that is the number of seconds that visitors stayed in front of the corresponding objects. All these objects were visited in a linear sequence, that is one after the next, with no repetitions or change of path. I will show in Chapter 7 how I use this data to train the parameters of the Bayesian network which drives the museum wearable experience.



For an example of a complete exhibit evaluation assessment see the report on the San Jose Tech Museum of Innovation Galleries by Randi Korn & Associates, Inc. [<http://www.randikorn.com/dwdocs/Summaries.html>].



Figures 18 and 19. Annotations of visitor's path and duration of stay at MIT Museum's Robots and Beyond Exhibit

Intro 1	Lisp 2	Minsky Arm 3	Robo Arm 4	Falcon 5	Phantom 6	Cogs Head 7	Quad- 8	Uniroo 9	Dext Arm 10	Kismet 11	Baby Doll 12	TYPE
0	5	5	0	13	0	10	0	0	0	0	0	busy
0	0	20	0	30	40	0	0	30	5	24	0	slctv
0	0	0	0	10	0	75	0	0	0	0	10	slctv
0	0	20	0	20	130	10	55	82	25	0	5	slctv
0	0	15	10	10	5	0	0	0	0	0	0	busy
0	0	5	5	5	0	3	0	0	70	0	0	busy
0	0	33	0	60	17	0	0	0	16	0	13	slctv
0	38	10	13	38	10	21	0	0	18	0	43	slctv
0	0	30	0	0	10	10	0	0	5	0	0	busy
0	6	40	15	25	40	0	82	82	34	30	18	greedy
0	0	0	0	35	15	10	0	15	55	20	10	slctv
0	31	45	15	25	10	5	0	0	0	40	0	slctv
0	0	15	15	27	20	0	0	0	35	0	3	slctv
0	18	0	0	30	41	0	0	0	23	15	50	slctv
141	0	0	0	0	0	0	0	0	0	0	110	slctv
10	23	20	0	25	40	26	56	56	0	7	20	greedy
5	18	0	0	3	0	0	0	0	10	55	24	slctv
140	45	30	0	0	0	0	0	0	0	0	0	slctv
0	1	8	0	0	0	0	0	0	5	12	10	busy
3	0	0	0	20	19	0	0	51	0	0	0	slctv
30	15	0	5	0	0	3	0	0	0	0	0	busy
5	38	0	0	140	35	0	0	0	0	25	0	slctv
15	0	0	0	10	5	10	3	5	10	0	10	busy
3	20	10	22	0	0	0	25	15	60	0	30	slctv
3	20	10	22	0	0	10	0	0	0	0	0	slctv
180	2	0	0	20	15	25	0	0	0	0	40	slctv
2	0	0	0	0	0	0	0	39	0	10	0	busy
3	35	5	5	0	20	0	0	0	0	0	10	slctv
3	0	10	0	0	0	5	0	0	0	0	0	busy
3	35	5	5	0	0	0	15	10	3	5	15	busy
15	0	37	0	0	0	5	0	0	0	0	0	busy
15	7	0	0	33	15	0	0	0	0	55	0	slctv
15	0	0	0	3	0	0	0	0	0	55	0	busy
10	30	3	0	0	0	15	0	0	5	0	35	slctv
0	43	10	0	47	0	20	55	55	35	0	0	slctv
3	0	0	0	3	0	0	0	0	0	0	15	busy
3	0	0	0	74	17	96	0	0	0	0	0	slctv
6	0	0	0	4	0	17	0	0	0	0	0	busy
3	41	10	0	20	9	0	14	8	10	31	0	busy
3	23	5	0	20	3	20	5	10	5	5	0	busy
5	10	5	0	0	10	0	0	0	10	65	15	busy
5	0	35	0	6	0	7	0	5	35	40	6	busy
3	60	15	30	40	30	10	0	0	0	0	0	slctv
10	45	45	60	35	0	0	0	5	10	20	0	slctv
3	27	50	39	30	0	0	0	15	20	0	0	slctv

Table 1. Visitor Tracking data limited to 12 selected objects at MIT Museum's *Robots and Beyond* Exhibit

### **3.6. The Museum Wearable: demonstration prototype**

To monitor the visitor's behavior in the museum, and deliver a story as a function of the visitor's evolving path, I have built a first physical implementation of the museum wearable using uniquely a location sensor. The location sensor informs the wearable on the wearer's location in the exhibit, and proximity to an object on display. The location system is made by a network of tiny infrared devices which transmit a location identification code to the receiver worn by the user and attached to the display glasses. The transmitters have the size of a 9V battery, and are placed inside the museum, next to the regular museum lights. They are built around a PIC microcontroller and their signal can be detected as far as about 30 feet away within a cone range of approximately ten to thirty degrees (Chapter 6).

While using only one sensor, may seem like a limiting factor in constructing an interactive experience such as the one described, having such long range infrared location identification sensor can provide a great deal of useful information for the targeted application. With the location identification receiver, connected to the wearable through the serial port, the museum wearable can measure a sampled path of the visitor throughout the exhibit, including how long the visitor stays in proximity of the tagged object on display, and his/her overall strategy of exploration.

The sto(ry)chastics approach (Chapters 4 and 5), allows the author to model the sensor, the visitor, and the content selection mechanism with a Bayesian network.

A simulation of how a more accurate estimation of the visitor's interests could be achieved using multiple sensors is provided in Section 5.3.

The following Chapter provides an introduction to Bayesian networks and explains how they can be used to estimate the visitor's type uniquely using the information provided by the infrared location identification sensor.

Chapter 5 illustrates the content selection mechanism, which edits in real time for the visitor a story about the object on display, from pre edited short segments, such that the overall edited story best matches the visitor's type and interests.



## Chapter 4

# Estimating the visitor's intentions with Bayesian networks

### 4.1. Bayesian Networks

Over the last decade, a method of reasoning using probabilities, variously called belief networks, Bayesian networks, probabilistic causal networks, influence diagrams, knowledge maps, constraint networks, qualitative Markov networks, and so on, has become popular within the AI probability and uncertainty community, and more recently, the machine learning, and pattern recognition communities. In the remaining part of this document I will use the term Bayesian networks, as it is the more widespread throughout the above mentioned research communities. The term Bayesian networks unfortunately suggests inappropriate comparison to neural networks. Bayesian networks should not be confused with neural network: they are more closely related to expert systems than to neural networks, and do not necessarily involve learning. The fundamental difference between the two types of networks is that a neural network in the hidden layers does not in itself have an interpretation in the domain of the system, whereas all the nodes of a Bayesian network represent concepts that are well defined with respect to the domain. The construction of a Bayesian network requires detailed knowledge of the domain in question. It is true that to construct a Bayesian network one needs to know many probabilities in advance. However, there is not a considerable difference between this number and the number of weights and thresholds that are needed to specify a neural network, and these can only be learned by training. In neural networks it is therefore impossible to utilize domain knowledge that one may have in advance. Probabilities for Bayesian networks, can instead be assessed using a combination of theoretical insight, empiric studies independent of the constructed system, training, and subjective estimates. Finally it should be mentioned that in the construction of a neural network the route of inference is fixed. It is decided in advance, about which relations information is gathered, and which relations the system is expected to calculate. Bayesian networks are much more flexible in that respect.

A Bayesian network is a graphical model which encodes probabilistic relationships amongst variables of interest. Such graphs not only provide an attractive means for modeling and communicating complex structures, but also form the basis for efficient algorithms, both for propagating evidence and for learning about parameters. Bayesian networks encode qualitative influences between variables in addition to the numerical parameters of the probability distribution. As such they provide an ideal form for

combining prior knowledge [Heckerman, 1999], which might be limited solely to experience of the influences between some of the variables of interest, and data. When used in conjunction with statistical techniques, Bayesian networks have several advantages for data analysis. One, because the model encodes dependencies among all variables, it readily handles situations where some data entries are missing. Two, a Bayesian network can be used to learn causal relationships, and hence can be used to gain understanding about a problem domain and to predict the consequences of intervention. Three, because the model has both causal and probabilistic semantics, it provides a representation for combining prior knowledge and data.

Bayesian networks are a specific type of graphical model. A Bayesian network is a specific type of graphical model called a *directed acyclic graph* (DAG) (as opposed to a UG: Undirected Graph) [Neapolitan, 1990]. That is, all of the edges in the graph are *directed* and there are *no cycles*. The nodes represent random variables that are variables of interest (i.e. the temperature of a device, a feature of an object, the occurrence of an event), and the links represent causal influences among the variables. The strength of an influence is represented by conditional probabilities that are attached to each cluster of parent-child nodes in the network [figure 20].

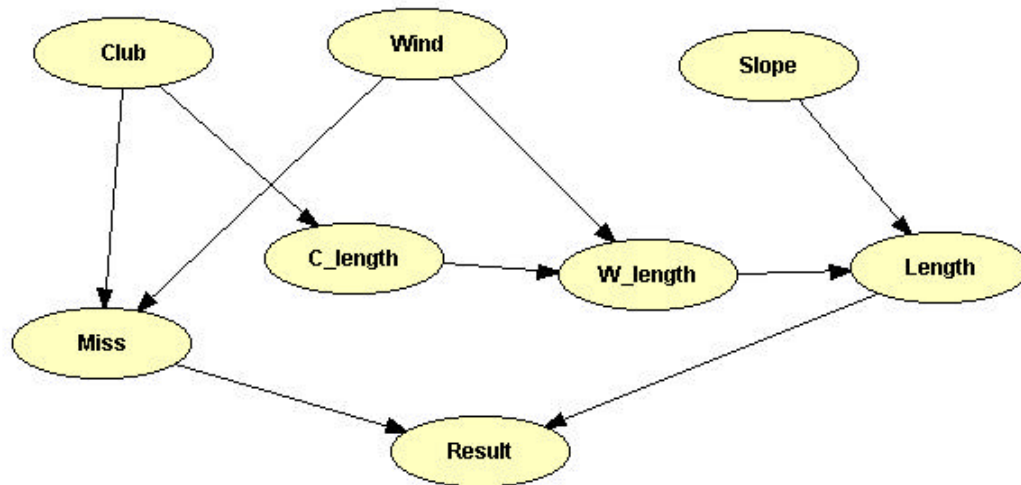


Figure 20. Example of a Bayesian network which models a golf game.

With the structure of the graphical model, the Bayesian network expresses a set of conditional independence relations amongst the variables of the network. Any information or computation on any variable of the model, such as calculation of posterior probabilities, requires knowledge of the joint probability distribution of all variables in the network. The knowledge of independence relations amongst the variables of the probabilistic model is important to be able to find a simpler factorization for the joint distribution of the variables. The graphical structure of the Bayesian network encodes all independence relations amongst the variables of the network in an easily readable and

intuitive way. More specifically, if  $\mathbf{U} = \{X_1, X_2, \dots, X_N\}$  is a set of  $N$  random variables, let  $p(\mathbf{U})$  represent the joint distribution for  $\mathbf{U}$ . It is well known that for moderately large  $N$ , specification and manipulation of  $p(\mathbf{U})$  directly is intractable unless there exist considerable structure in the probability model. For example with  $N$  binary variables, a model with no independence structure requires the specification of  $O(2^N)$  probability values. Furthermore, calculations of particular posterior probabilities given observed evidence will also tend to scale exponentially in  $N$ , rendering such models useless in practice. This intractability has been well-known in different disciplines for some time and there has been considerable, and often independent work in different areas to exploit independence structure to achieve tractability [Smyth, 1998]. The conditional independence relations in a Bayesian network are represented by the *missing edges* of the graphical model. If variable  $X_i$  does not depend directly on variable  $X_j$ , then there is no edge between them, as a node is connected only to those other nodes on which it directly depends. The parameters of the network consist of the specification of the joint probability distribution  $p(\mathbf{U})$ . This specification is in a factored form, and the factors are defined locally on the nodes of the graph.

The conditional independence relationships between the variables of a Bayesian network is described by the property of *d-separation*, hence can be immediately read off the graph [figures 21,22]. Two network variables A and B are d-separated if all paths between them are blocked [Minka, 1999]. A path between A and B is blocked if there is a node C such that:

1. the path has converging arrows at C and none of C or its descendants are given
2. the path does not have converging arrows at C and C is given.

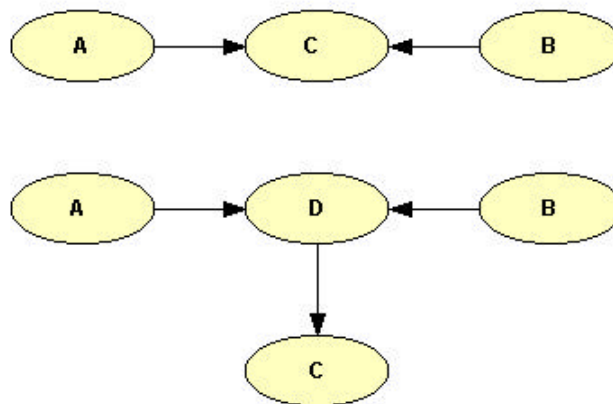


Figure 21. The path has converging arrows at C and none of C or its descendants are given

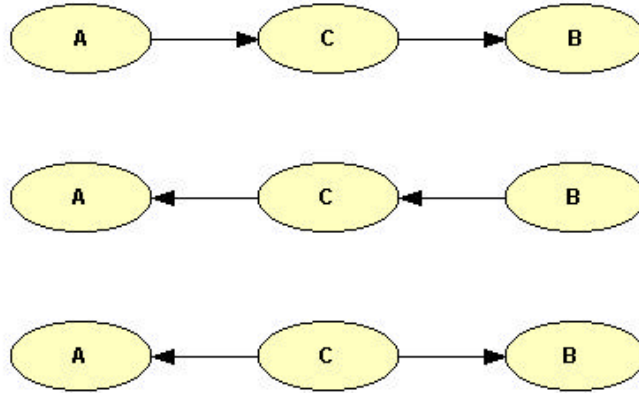


Figure 22. The path does not have converging arrows at C and C is given

An example of how effectively the joint probability distribution can be easily factorized by reading conditional independence relations amongst the variable of the network from the property of d-separation of the network is given in figure 23 [from Minka, 1999]:

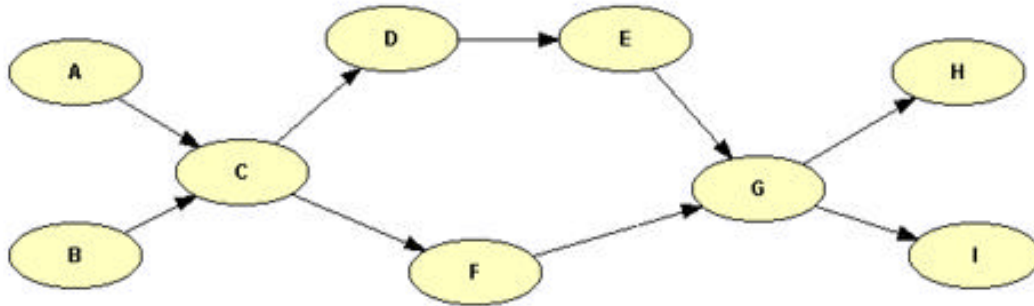


Figure 23. Example of simplification of joint distribution obtained by looking at the d-separation properties of the graph

Mathematically this graph tells us that the joint distribution can be factored as:

$$p(U) = p(A, B, C, D, E, F, G, H, I) =$$

$$p(A)p(B)p(C | A, B)p(D | C)p(E | D)p(F | C)p(G | E, F)p(H | G)p(I | G)$$

Jensen [Jensen, 1996] describes two introductory examples that illustrate essential points to consider when reasoning with uncertainty with Bayesian network. These examples model respectively *conditional independence* and *explaining away*.

**Example 1.** Crowded Museum: conditional independence  
(modified from Jensen's "Icy Roads" example)

Frank, an ambitious artist, is impatiently awaiting the arrival of Henry and Will, two well known museum curators with whom he needs to discuss placement of his artwork in the museum galleries they direct. Frank is quite agitated as they are late, and he has a third appointment lined up in the afternoon. Wondering why they could be so late, he recalls that they told him, that before the meeting, they were planning to visit the Impressionists



exhibit that was being shown in town. Frank seems to recall that today is also the last day of the exhibit, and therefore the museum they're visiting might be very crowded. Both Henry and Will are meticulous and diligent art critics, and therefore if the museum is crowded, they may be there for a very long time. At that point, Liz, Frank's friend, comes into the room and tells him that Will called saying that he would be late as he was busy at the Impressionists' gallery. "Ah, I knew it" – Frank replies – "today is the last day of that exhibit and it must be terribly crowded. Therefore Henry is probably stuck there as well. I'd better head on to my next meeting". "Last day?" – Julie replies – "It's far from being the last day, the exhibit has been extended for another two weeks, and furthermore the museum has special visiting hours for art critics". Frank is relieved. "Too bad for Will, he must have met one of his sponsors and has not been able to disengage to come here on time. Let's give Henry another half hour".

This example can be formalized as follows. We have three variables: crowded galleries (C), Henry is late (H), and Will is late (W). Let all events be represented by variables with two states: "yes" and "no." To each event is associated a certainty, which is a real number. C has the effect of increasing certainty of both H and W. For quantitative modeling we need initially  $p(C)$ ,  $p(H | I)$ , and  $p(W | I)$ . This reflects the fact that only knowledge of Crowd is relevant for Will and Henry being late, in this model [figure 24].

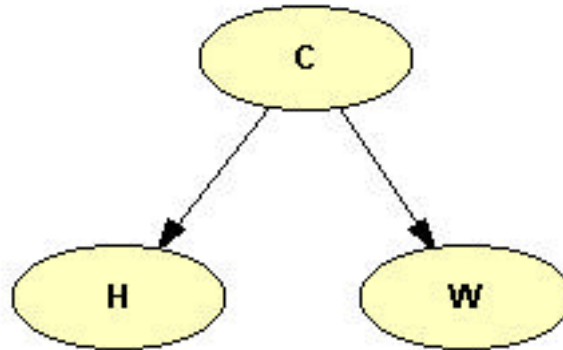


Figure 24. Crowded museum Bayesian network example

The initial probability assignment for this model is given in table 2. Initially Frank believes that the Impressionists' museum is crowded, that is  $p(C=\text{yes})=0.7$ . The other conditional probability tables reflect the fact that if the museum is crowded, Henry and Will are likely to be late:  $p(H=\text{yes} | C=\text{yes})=0.8$ , otherwise they are usually on time:  $p(H=\text{yes} | C=\text{no})=0.1$ .

		$p(H C)$	C: yes	C: no	$p(W C)$	C: yes	C: no
C: yes	0.7	H: yes	0.8	0.1	W: yes	0.8	0.1
C: no	0.3	H: no	0.2	0.9	W: no	0.2	0.9

Table 2. Crowded museum Bayesian network: initial probabilities.

Therefore, numerically, the network initially is as shown in figure 25:

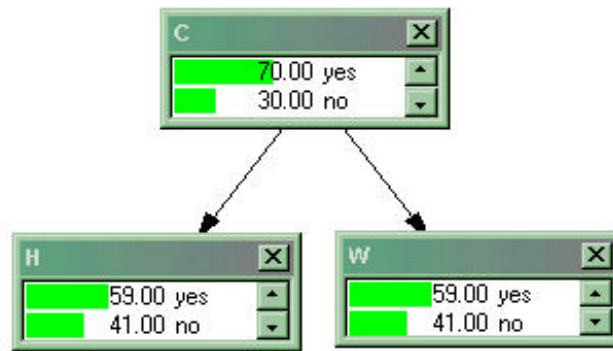


Figure 25. Crowded museum Bayesian network initial probabilities

When Frank is told that Will is stuck at the museum and is going to be late, he is doing a reasoning in the opposite direction to the causal arrows. He gets an increased certainty of C [figure 26]. The increased certainty of C in turn creates a new expectation, namely an increased certainty of H.

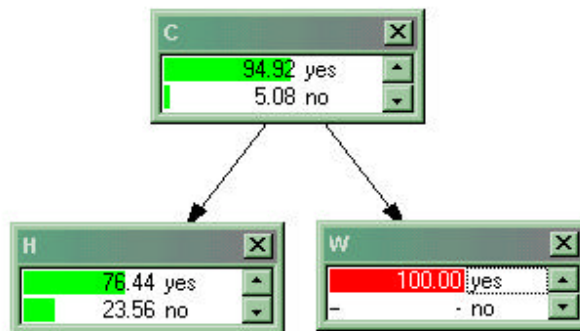


Figure 26. Crowded museum Bayesian network: updated probabilities after evidence about Will being late is introduced.

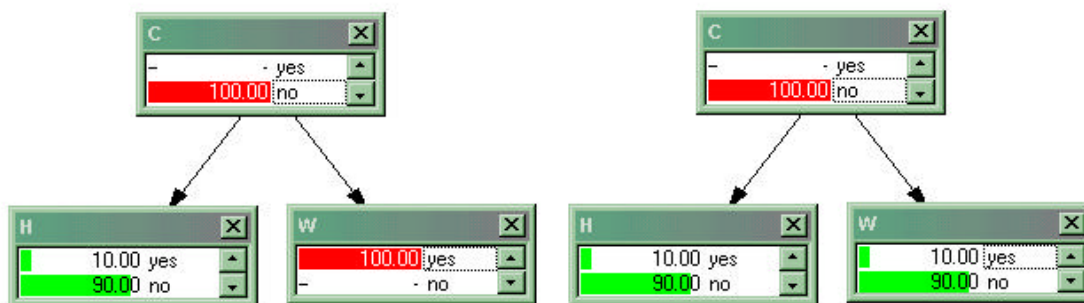


Figure 27. Crowded museum Bayesian network: after evidence on Crowd is introduced, Will's being late for the meeting has no influence on Henry: the probability of H is the same, independently of the probability value of W.

Next, when Liz tells him that the Impressionists' museum cannot possibly be crowded, the fact that Will is stuck at the museum cannot change his expectation about the museum being crowded, and consequently, Will's being late for the meeting has no influence on Henry [figure 27].

This is an example of how dependence/independence changes with the information at hand. When nothing is known about how crowded the Impressionists' exhibit is, then H and W are dependent: information on either event affects the certainty of the other. However when the crowd information is known for certain, then they are independent: information on W has no effect on the certainty of H and viceversa. This phenomenon is called conditional independence.

**Example 2.** Interactive Museum: Explaining away  
(modified from Jensen's "Wet Grass" example)

Henry is the director of the famous Come-See-Touch interactive science museum in Los Angeles. One morning, as Henry is leaving the main gallery, he notices that the "Highlights in the history of science" interactive demonstration is not working. It is a complex interactive piece which requires a daily initialization procedure. Is it not working because somebody forgot to run its daily initialization, or is it just due to the intermittent power outages that they've had in Los Angeles all summer ? His belief in both events increases. Next he notices that the neighbor interactive demonstration, the "Wonders of Nature" interactive table, is also not working. Given that the table is simply "plug-and-play" and does not require special assistance, he is now almost certain that the galleries have been subject to yet another power outage.

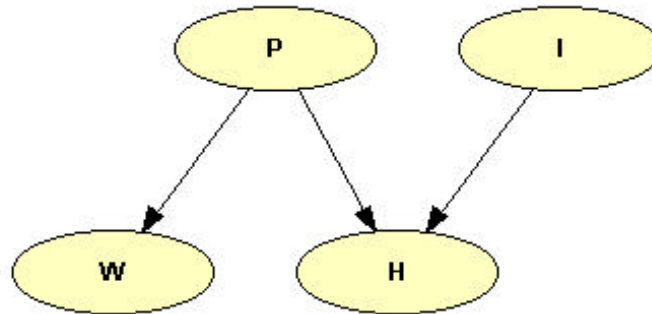


Figure 28. Interactive museum Bayesian network

In this example we have four variables: power outage (P), forgot initialization (I), "Wonders of Nature" demonstration down (W), and "Highlights in the history of science" demonstration down (H) [figure 28]. They all have two states: "yes" and "no." The initial probabilities are given in table 3:

								I:yes		I:no	
				$p(W P)$	P: yes	P: no	$p(H P,I)$	P: yes	P: no	P: yes	P: no
P: yes	0.2	I: yes	0.1	W: yes	1	0.2	H: yes	1	0.9	1	0
P: no	0.8	I: no	0.9	W: no	0	0.8	H: no	0	0.1	0	1

Table 3. Interactive Museum Bayesian Network: initial probabilities.

When Henry notices that the “Highlights in the history of science” demonstration is not working, his certainty of both P (power outage) and I (forgot initialization procedure) increases. The increased certainty of P in turn, creates an increased certainty of W [Figures 29, 30].

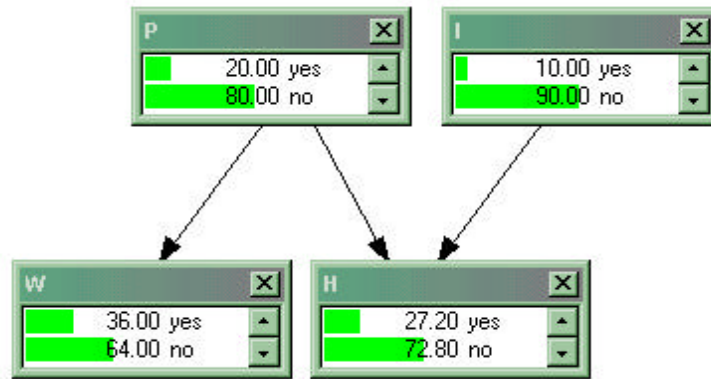


Figure 29. Interactive museum Bayesian network: initial probabilities.

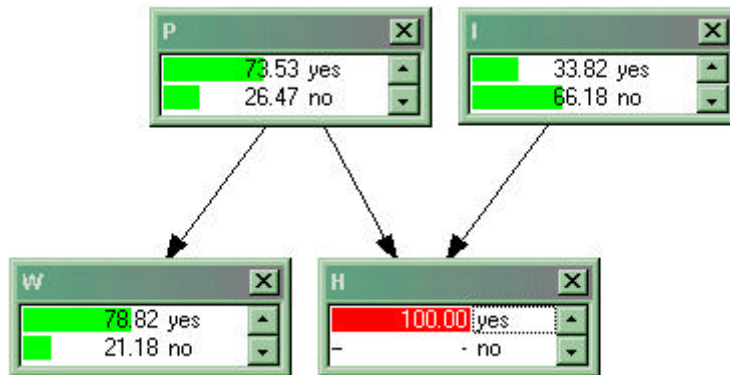


Figure 30. Interactive museum Bayesian network: probabilities after H=yes is introduced as evidence.

Then Henry checks the Wonders of Nature demonstration, and when he discovers that is also down, he immediately increases certainty of P (power outage) [figure 31].

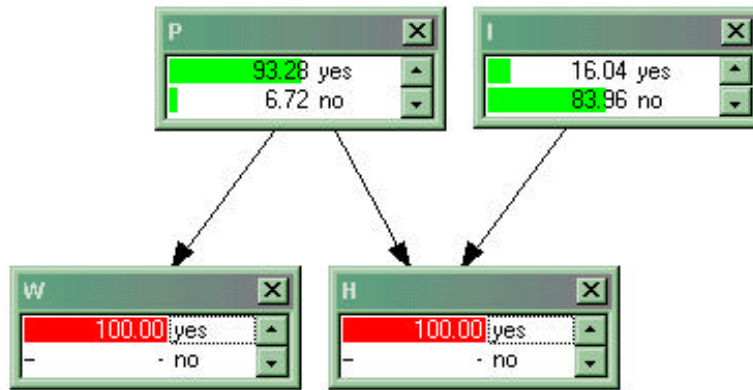


Figure 31. Interactive museum Bayesian network: probabilities after H=yes and W=yes are introduced as evidence.

This reasoning step is hard for machines but natural for humans, and it's called explaining away: the Highlights demonstration being down has been explained, and thus there is no longer any reason to believe that the initialization procedure has been forgotten. Hence the certainty of I is reduced. The reason why the probability of I (forgot initialization procedure) does not drop to the prior probability of 0.1 is that the Wonders of Nature demonstration may be down for other reasons (it could be broken). This is reflected in the probability  $p(W=\text{yes} \mid P=\text{no})=0.2$ .

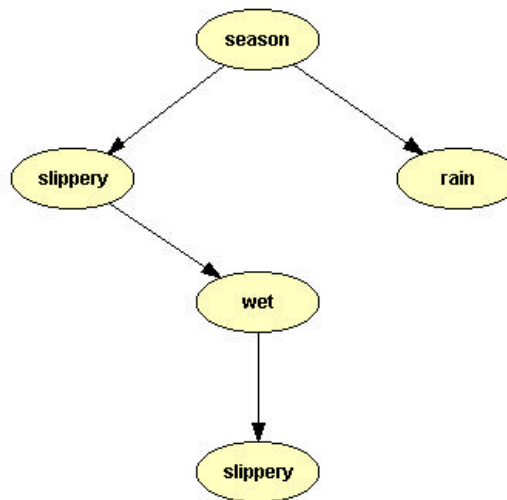
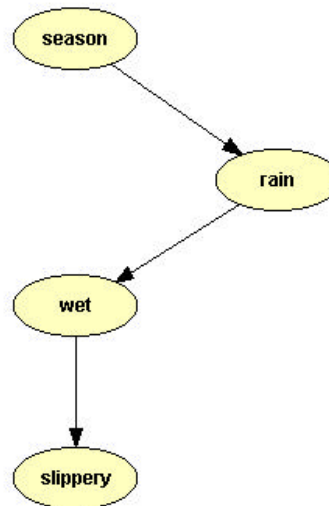
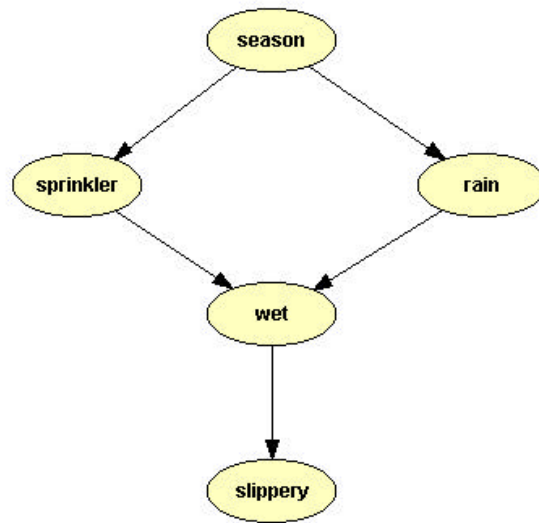
The numeric values shown after introducing evidence, and reported in the figures illustrating the above examples, are the result of inference, or probability update for the network. This is the operation of obtaining revised belief for some or all the nodes in the domain, after evidence has been introduced. One can think of a propagation as the computation of certain marginals of the full joint probability distribution over all variables. As is well-known, the distribution of an individual variable can be found by summing/integrating out all other variables of this joint probability distribution. More efficiently in the above examples they are computed using a two-pass propagation operation called Collect Evidence and Distribute Evidence, which can be thought of as a parametrization of the marginalization method [Cowell et al., 1999]. This method is implemented in HUGIN [Andersen, 1989], and HUGIN is the software library used to calculate inference for these examples.

Pearl observes that a Bayesian network constitutes a model of the environment rather than, as in many other knowledge representation schemes (i.e. logic, rule-based systems, and neural networks), a model of the reasoning process [Pearl, 1988]. One could argue that in some cases knowledge equals the environment, and in others the network represents abstract ideas. A Bayesian network allows the investigator to answer a variety of queries, including: associational queries, such as "Having observed A, what can we expect of B?"; abductive queries, such as "What is the most plausible explanation for a given set of observations?"; and control queries, such as "What will happen if we intervene and act on the environment?". Answers to the first type of query depend only on probabilistic knowledge of the domain, while answers to the second and third types rely on the causal knowledge embedded in the network. Both types of knowledge,

associative and causal, can effectively be represented and processed in Bayesian networks. The associative facility of Bayesian networks may be used to model cognitive tasks such as object recognition, reading comprehension, and temporal projections. For such tasks, the probabilistic basis of Bayesian networks offers a coherent semantics for coordinating top-down and bottom-up inferences, thus bridging information from high level concepts and low level percepts [Pearl, 1988]. The most distinctive feature of Bayesian networks is their ability to represent and respond to changing configurations. Any local reconfiguration of the mechanisms in the environment can be translated, with only minor modification, into an isomorphic reconfiguration of the network topology. In the network given as example below, to represent a disabled sprinkler, we simply delete from the network all links incident to the node “Sprinkler”. To represent a pavement covered by a tent, we simply delete the link between “Rain” and “Wet. This flexibility is often cited as one of the main advantages of Bayesian networks, as it enables them to manage novel situations instantaneously, without requiring training or adaptation [figures 32,33,34].

The examples above also provide examples of the basic possible connections between the nodes of a Bayesian network. If A, B, C are nodes, the three main connection types are [figure 35]:

- serial connection: A has an influence on B which in turn has an influence of C. Evidence on A will influence the certainty on B, which then influences the certainty on C. Similarly, evidence on C will influence the certainty on A through B. On the other hand, if the state of B is known, then the channel is blocked, and A and C become independent as evidence is transmitted through a serial connection unless the state of the variable in the connection is known.
- converging connection: if nothing is known about A except what may be inferred from knowledge of its parents, then the parents are independent: evidence on one of them has no influence on the certainty of the others. If evidence is given for A, then the parents become dependent due to the principle of explaining away (conditional dependence).
- diverging connection: this is a generalization of the crowded museum example: influence can pass between all the children of A unless the state of A is known (conditional independence).



Figures 32, 33, 34. Flexibility of Bayesian networks

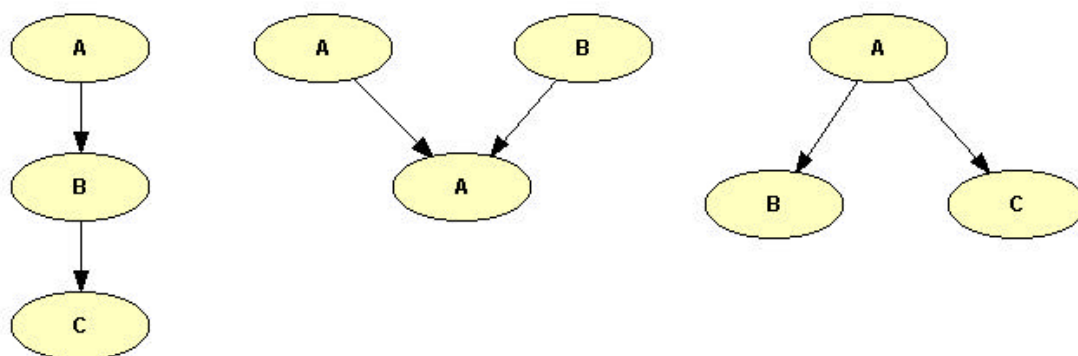


Figure 35. Basic connection types for Bayesian networks: serial, converging, and diverging connections.

Another observation which can be made from the above examples is that by using graphs, not only it becomes easy to encode the probability independence relations amongst variables of the network, but it is also easy to communicate and explain what the network attempts to model. Graphs are a natural medium for representing information in a compact form which humans can grasp, understand, and use. Whittaker [Whittaker, 1990] provides a number of examples which clearly demonstrate that even with relatively few variables it is much easier to reason about independence relations using a graph than it is without. In addition, the fact that the graphical model forces the modeller to explicitly encode independence assumptions can be extremely useful in model-building [Smyth, 1998].

A Bayesian network is called an influence diagram [Howard and Matheson, 1981] when working with decision making. In decision theory one also specifies the desirability of various outcomes, i.e. their utility, and the cost of various actions that might be performed to affect the outcomes. The idea is to find the action (or plan) that maximizes the expected utility minus costs. An influence diagram includes action nodes, i.e. nodes indicating actions that can be performed, and utility nodes, i.e. nodes indicating the values of various outcomes.

Various operations are typically performed on Bayesian networks. Inference is the problem of calculating posterior probabilities for variables of interest given observed data and given a specification of the probabilistic model. Typical inference problems include calculating the probability of a class variable given observed features (in classification problems) and calculating the probability of observed data under various different models (as in speech recognition). The related task of maximum a posteriori (MAP) identification is the determination of the most likely state of a set of unobserved variables, given observed data and the probabilistic model. The learning or estimation problem is that of determining the parameters, and/or the structure of the probabilistic model from the data.

Following are some guidelines on some of the techniques typically used in Bayesian networks for inference and learning.



#### 4.1.1. Inference in Bayesian Networks

Inference is the task of computing the probability of each value of a node in a Bayesian network when other variables' values are known. This is called equivalently probability updating, network updating, or inference. In general this computation is NP-hard [Cooper, 1987]. Depending on the particular structure of the network, the algorithm used, and the accuracy of implementation, networks as small as a dozen of nodes can take too long, or networks in the thousands of nodes can be done in acceptable time. The tradeoffs are between speed, complexity, generality, and accuracy. Given that getting an exact solution is NP-hard, researchers have developed a variety of methods to obtain approximate solutions, which with high probability are within some small distance of the correct answer.

One class of networks which can be updated with exact methods is that of singly connected networks. A singly connected network, also called polytree, is one in which the underlying undirected graph has no more than one path between any two nodes. The underlying undirected graph is obtained by ignoring the direction of the edges between nodes in the network. A version of probability updating in singly connected networks through message passing was presented by Kim and Pearl [Kim and Pearl, 1983]. To evaluate multiply connected networks exactly, one needs to transform the network into an equivalently singly connected one. Pearl describes in depth this technique, called belief update or belief propagation in [Pearl, 1988, chapter 4]. A popular, fast, technique to do this is one by Lauritzen and Spiegelhalter [Lauritzen and Spiegelhalter, 1988] and later improved by Jensen [Jensen, 1990], also known as the Hugin method. This is the technique used in the numeric examples above. An alternative to this probability update method is the one proposed by Shafer and Shenoy [Shafer and Shenoy, 1990]. A technique called lazy propagation proposed by Madsen and Jensen [Madsen and Jensen, 1999] merges the Shafer-Shenoy and Hugin propagation. Cowell [Cowell, 1999] introduces a more general method of inference in Bayesian networks, called the junction tree algorithm. Approximate techniques come in various flavors depending on the nature of the network. In essence they randomly set values for some of the nodes, and then use these to select values for the other nodes. The statistics of the answers on the values of each node give the final update value. Neal [Neal, 1993] introduced Markov Chain Monte Carlo (MCMC) methods for Bayesian networks, inspired by research in statistical physics. Gilks et al. [Gilks et al., 1994] have developed a quite popular and effective system, called BUGS, for Gibbs sampling in Bayesian networks. In contrast to the approach that applies probability propagation to multiply connected trees, and the computationally intensive stochastic approach of Monte Carlo, Jordan et al. [Jordan, Ghahramani, Jaakkola, Saul, 1999] propose an elegant variational formulation of probability updating. Loopy belief propagation is a technique by Weiss et al. [Weiss, 1997] which entails applying Pearl's algorithm to the original graph, even if it has loops (undirected cycles). In theory, this runs the risk of double counting, but in certain cases (e.g., a single loop), events are double counted equally, and hence cancel to give the right answer. Generalized Belief Propagation (GBP) is a very accurate but computationally expensive belief propagation algorithm, inspired by statistical physics, by Yedidia, Freeman, and Weiss [Yedidia, Freeman, and Weiss, 2000]. Recently Minka [Minka, 2001] proposed a new fast probability update approximation technique called

“Expectation Propagation”, which greatly reduces computational expense with respect to all previous techniques.

Typically there is no single probability update technique, approximate or exact, which works well for all kinds of networks, and researchers need to choose the algorithm which best suits their network and problem, amongst the available ones.

#### **4.1.2. Learning in Bayesian Networks**

Learning in Bayesian networks refers to learning either the topology or the conditional probability distributions (parameters) of the network. Depending on the problem at hand, either or both of these may be pre-defined by the expert, or domain knowledge about the problem, or may be learned. Heckerman [Heckerman, 1999] provides a good introduction to learning with Bayesian networks.

If the structure of the network is known, Expectation Maximization (EM) is the Maximum Likelihood approach used to learn the parameters [Lauritzen, 1995]. Zweig [Zweig, 1998] discusses how to use EM in discrete Bayesian networks for the purpose of doing speech recognition. Jordan et al. [Jordan, Ghahramani, Jaakkola, Saul, 1999] introduce variational methods for learning and inference in graphical models. Another class of learning methods is based on Monte-Carlo or sampling methods [Neil, 1993].

If the structure of the network is not known, various techniques are available to learn the network structure from data. This is an active field of research, especially for hybrid networks containing both continuous and discrete nodes. Cooper and Herkovits [Cooper and Herkovits, 1992] present the K2 algorithm to find the most probable belief network structure, from a database of cases. Chickering, Heckerman and Meek [Chickering, 1997] describe a Bayesian approach to structure learning with a greedy algorithm that searches both for global and local structures simultaneously. Sprites, Glymour, and Scheines [Sprites, 2000] proposed a constraint-based approach to structure learning called the PC algorithm, which is now implemented in the Hugin library. Friedman [Friedman, 1997] explains a method for learning both the network structure and the data, which in a later paper, he renames the structural EM algorithm (SEM) [Friedman, 1998].

If more than one network topology is available, such as for example, one given by the expert and one learned, the model selection problem arises. One method for evaluating a potential structure is to compute the joint probability for the data and structure  $p(D, S)$ . Using Bayes' theorem, this breaks down into computing the posterior probability (the likelihood) and the prior probability of the structure of the data (relative posterior probability) [Heckerman, 1999]. Similarly, it is possible to choose between network topologies using the “evidence” for the model [Mackay 1992a, 1992b, 1995]. The evidence framework has undergone some discussion and controversy in the Bayesian network community [Wolpert, 1993], yet it stays as a good first guideline for model selection. Minka [Minka, 2001] proposes expectation maximization for model selection. Neal [Neal, 1993] shows the use of Markov Chain Monte Carlo (MCMC) for model comparison.

### 4.1.3. Dynamic Bayesian Networks

Dynamic Bayesian Networks (DBN) model those problems where the environment evolves over time. The model assumes that the Markov property holds, i.e. the current state is affected only by the last state, and the action taken at that state. Based on the observed history, the model evaluates the posteriors of a given state at a given time. It typically answers queries related to how the system will evolve (forecasting or prediction). A Dynamic Bayesian Network is defined by a prior network and a transition network [figure 36]. The transition network illustrates, for all time slices, what the probabilities are for each variable, conditioned of the other variables, from the previous and the same time slice. A DBN needs to satisfy the following conditions: (a) it has the same structure at any time slice  $t$  and (b) the only transition (cross-slice) edges allowed are those that extend from slice  $t$  to slice  $t+1$ .

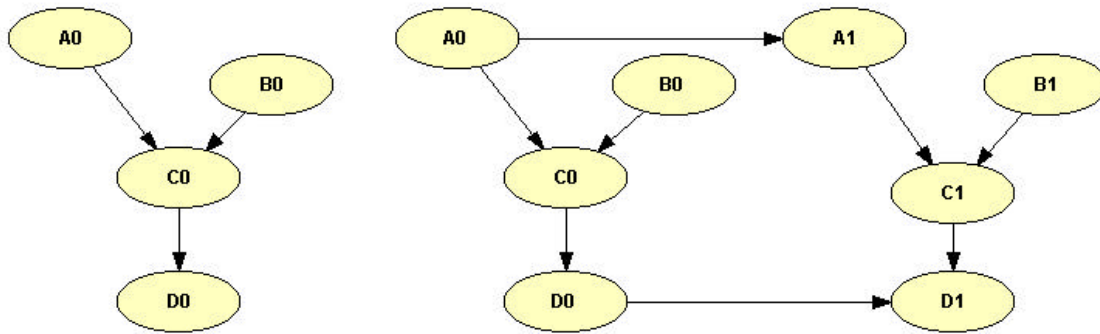


Figure 36. Prior network and transition network for a Dynamic Bayesian Network.

The set of variables and probability definitions are the same for each time slice, with the exception of the prior network, in the initial time slice, which has its own probability distribution. Given the prior and transition network one can construct a dynamic Bayesian network of arbitrary length [figure 37]. A broad corpus of exact and approximate inference and learning techniques from the BN literature can be applied to dynamic Bayesian networks [Ghahramani, 1997].

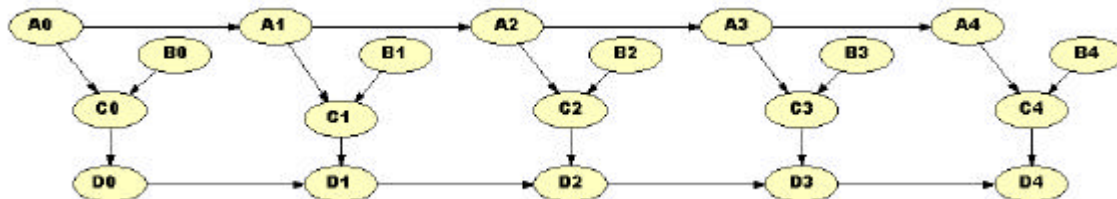


Figure 37. The same dynamic Bayesian network with 5 time slices.

Note that it is possible to model a dynamic Bayesian network with a limited number of time slices, using a “regular” Bayesian network, by replicating by hand the nodes for each slice and the transition links.

The merit of Dynamic Bayesian Networks is that they generalize widely diffused modeling techniques, such as HMMs and Kalman Filters. They capture a much richer structure including spatial and temporal multiresolution structure, distributed hidden state representations, and multiple switching linear regimes [Ghahramani, 1997]. For this reason they have been successfully applied to a variety of tasks such as multimodal sensing for gesture recognition [Pavlovic, 1999], speech recognition [Zweig, 1998], traffic surveillance [Forbes, 1995], and body motion understanding [Pavlovic, Rehg, Cham, and Murphy, 1999].

#### **4.1.4. Related modeling techniques: HMMs, CHMMs, Kalman Filters, and Markov Random Fields.**

Many time series models, including Hidden Markov models (HMMs), typically used in speech recognition, and Kalman filter models, used in filtering and control applications, can be viewed as examples or subsets of dynamic Bayesian networks [Ghahramani, 1997].

HMMs consist of states, possible transitions between states, and the probabilities that in a particular state, a particular observation is made. An observation can be any variable of interest, from a phoneme, to a handwritten character. HMMs are called hidden because the state of an HMM cannot in general be known by simply looking at the observation. They are Markov in the sense that the probability of observing an output depends only on the current state and not on previous states. By looking at the observations, using an algorithm known as the Viterbi Algorithm, one can compute an estimate of the probability that a particular instance or stream observed was generated by that HMM. Another problem usually solved with HMMs is to compute the most probable HMM that generated an observed data stream. These two problems, as similar to the inference and structure learning problem for Bayesian networks, discussed earlier. An HMM can effectively be represented as the simplest kind of DBN, which has one discrete or continuous observed node per slice [figure 38]. Specifically, in an HMM the sequence of observations  $\{Y_t\}$  is modeled by assuming that each observation depends on a discrete hidden state  $X_t$ , and that the sequences of hidden states are distributed according to a Markov process. The joint probability for the sequences of states and observations, can be factored exactly as for a Bayesian network with the structure shown in figure X:

$$p(\{X_t, Y_t\}) = p(X_1)p(Y_1 | X_1) \prod_{t=2}^T p(X_t | X_{t-1})p(Y_t | X_t)$$

Consequently the conditional independencies in an HMM can also be expressed graphically using dynamic Bayesian networks, and the parameter and structure learning methods can be considered as special cases of the ones discussed earlier for Bayesian networks [see Ghahramani, 1997].

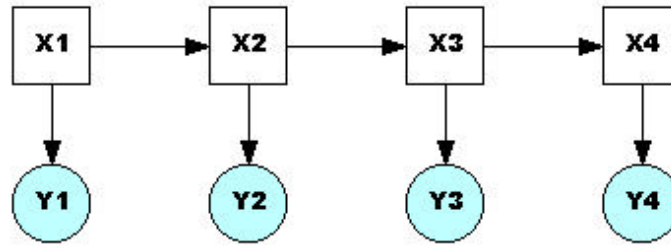


Figure 38. DBN representation of an HMM with 4 time slices.  
Circles denote continuous nodes, squares → discrete nodes, clear means hidden, filled cyan → observed.

Coupled HMMs extend HMMs to model two separate interacting processes, and have successfully been used for gesture recognition [Oliver, 1999], and freeway traffic modeling [Kwon and Murphy, 2000]. In this architecture two HMM chains are coupled via conditional probabilities modeling causal temporal influences between the hidden state variables [figure 39]. Basically, for each chain, the state at time  $t$  depends on the state at time  $t-1$  in both chains. Even though the topology of a coupled HMM resembles that of an ordinary HMM, the inference schemes of ordinary HMMs are not directly applicable to the coupled ones. Brand [Brand, 1996] describes a Viterbi-like approximation inference scheme for CHMMs. Pavlovic developed a variational inference approach [Pavlovic, 1999], using a DBN framework. Also within a DBN framework, Kwon and Murphy present approximate inference algorithms based on particle filtering (sequential Monte Carlo) and on the Boyen-Koller algorithm.

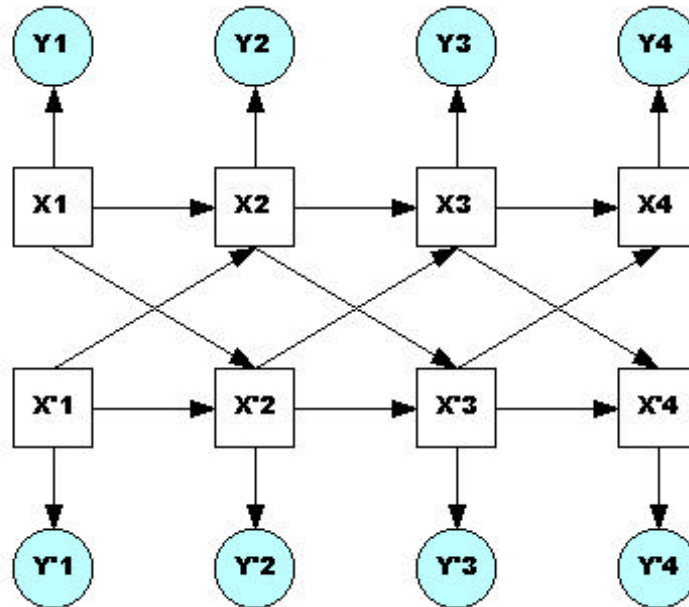


Figure 39. DBN representation of a coupled HMM.

A Linear Dynamical System, or state-space model, is a partially observed stochastic process with linear dynamics and linear observations, both subject to gaussian noise. The Kalman filter is an algorithm to perform filtering or prediction on this model. A Kalman filter performs least-squares optimal recursive estimation, and it can be described also as a graphical model. In state-space models, a sequence of real valued observation vectors  $\{Y_1, \dots, Y_T\}$  is modeled by assuming that at each time step  $t$ ,  $Y_t$  was generated from a real valued hidden state variable  $X_t$ , and that the sequence of  $X_s$  define a first order Markov process. If  $\{Y_t\}$  denotes sequences from  $t = 1$  to  $t = T$  then:

$$p(\{X_t, Y_t\}) = p(X_1)p(Y_1|X_1)\prod_{t=2}^T p(X_t|X_{t-1})p(Y_t|X_t)$$

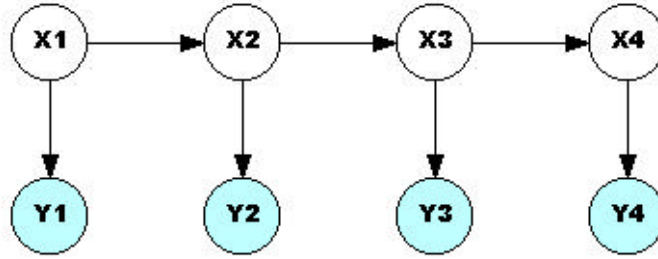


Figure 40. DBN representation of a Kalman filter with 4 time slices.

Circles denote continuous nodes, squares  $\rightarrow$  discrete nodes, clear means hidden, filled cyan  $\rightarrow$  observed.

Therefore, a Kalman filter has the same DBN topology as an HMM, where all the nodes are assumed to have linear-Gaussian distributions [figure 40].

Markov Random Field (MRF) [Kinderman and Snell, 1980] are two dimensional Markov chains represented by an undirected graph with nodes corresponding to variables [figure 41]. They were originally developed in statistical physics to model systems of particles interacting in a two dimensional or three dimensional lattice. Recently, they have been applied to problems in image analysis [German and German, 1984] where pixels play the role of particles in the physical system. The basic idea of a markov random field is that the conditional probability for a state variable  $s_{k,l}$  at position  $(k,l)$  in the field, is the same as the conditional probability of  $s_{k,l}$  given the states only in some local neighborhood. The joint distribution for an MRF can be written in a factorized form, as in the example below:

$$p(u, s, x, y) = p(u)p(s|u)p(x|u, s)p(y|x)$$

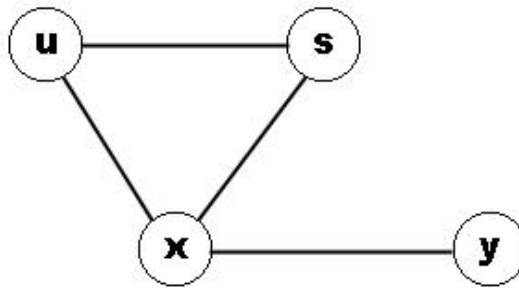


Figure 41. An example of Markov Random Field

It can be observed that an MRF is the UG (undirected graph) version of a Bayesian network, which is instead a DAG (directed acyclic graph). Since inference with an ADG is typically carried out by inference on a related UG, there is increasing interest in exploring the UG approximations for some applications involving ADGs [Saul and Jordan, 1996].

When some of the observed nodes are thought of as inputs (actions), and some as outputs (percepts), we can think of a DBN as a POMDP (Observable Markov Decision Process) [Murphy, 2000; Kaelbling, 1998].

From the discussion above it becomes clear that the formalism of HMMs, CHMMs, Kalman Filters, and Markov Random fields, can be expressed in the graphical framework of Bayesian Networks and Dynamic Bayesian Networks. The inference and learning algorithms known for Bayesian networks generalize previous algorithms used in the models described in this section. Therefore we can think of Bayesian networks as a superset of these previous techniques [Ghahramani, 1997; Smyth, 1998].

The ability to view seemingly different algorithms for different problems within a unified graphical model framework can provide powerful insights [Smyth, 1998]. More important is the fact the graphical model framework enables the construction and application of novel and relatively complex multivariate models in a straightforward and systematic manner [Pavlovic, 1999]. Specifically the Bayesian network formalism allows easy construction of hybrid models, which combine the use of HMMs and Kalman filters for example, as shown by Pavlovic and Rehg [Pavlovic, Rehg, Cham, and Murphy, 1999].

The flexibility of use of Bayesian networks is also illustrated by the wet floor example above. It become clear that especially in the modeling phase it is extremely useful to be able to add or delete nodes in the model without having to relearn or recompute all the conditional probability tables for the other nodes. This allows the system designer to have room for experimentation and trial, as it is possible to easily test multiple models, before selecting a final Bayesian graph that best models a problem, as shown in the next section.

Moreover, graphical models can easily be used for interpersonal communication. The graphical format is easy for humans to read, and it helps focus attention, for example in a group working together to build a model. In the context of this thesis for example, this allows the digital architect, or the engineer, to communicate on the same ground (the graph of the model) with the curator and therefore to be able to encapsulate the curator's domain knowledge, by getting appropriate feedback and applying multiple revisions to the model. Jensen [Jensen, 2001] sees Bayesian networks as a language, specifically a context-free language with a single context-sensitive aspect (no directed cycles). As a language, Bayesian networks are well defined and their formalism can easily be communicated to a computer.

For all these reasons, I chose to use Bayesian networks to model visitor's types at the museum, rather than any of the techniques described in this chapter. By using a superset of HMMs and CHMMs, and taking advantage of the modeling flexibility of Bayesian network, in addition to the ease of communication with non technically savvy people, such as the museum curators, I was able to carry out the work described in the following section and chapters in a relatively short amount of time.

To speed up the work, at least in the initial phase, in addition to the above mentioned advantages, there exist today various software packages available to the researcher who wishes to investigate applications or novel techniques for Bayesian Networks. The main three well developed packages are:

- Hugin [Andersen, 1989] [<http://www.hugin.com>]
- Kevin Murphy's matlab Bayesian Net Toolbox (BNT) [Murphy, 2001] [<http://www.cs.berkeley.edu/~murphyk/Bayes/bnt.html>]
- BUGS [Gilks, Thomas, Spiegelhalter, 1994] [<http://www.mrc-bsu.cam.ac.uk/bugs/>]

For this thesis I used Hugin for calculation of inference, in simulation and in real time. I also used Kevin Murphy's matlab BNT to learn the network parameters as described in section 7.1.

## 4.2. Modeling visitors' intentions at MIT's Robots and Beyond Exhibit

The museum wearable uses a Bayesian network to provide a real time estimate of visitor typology. An analysis of museum visitors, (see section 3.3) has led to identify three main visitor types (or strategies): a **greedy** type, who wants to know and see as much as possible, and does not have a time constraint, a **busy** type who just wants to get an overview of the principal items in the exhibit, and see little of everything, and the **selective** type, who wants to see and know in depth only about a few preferred items. Our system attempts to classify the visitor behavior in the museum according to the above



user types, and to present content tailored to these three types according to the intentions and interests expressed by their estimated behavior in the museum space. The information upon which the visitor type is identified is given by the infrared location sensor described in section 6.3. It is of course possible and desirable to assign priors to the user types: people entering the exhibit could be asked if they wish to declare themselves as belonging to one of the above user types, or if they wish instead to explore the exhibit uncommitted to any exploring strategy. It is also possible for people to behave differently in different stages of their visit: somebody could for example behave like a busy type during their first visit, and later behave like a selective type on a second visit. If people are given the option to declare their visiting strategy (busy/greedy/selective) the system needs to be able to account for those cases in which visitors change their mind: having declared themselves as belonging to a busy type for example at the beginning of their visit, they are later on carried away by what there are seeing and behave like a selective type. Finally, priors could be assigned to the different user types by having visitors spend some time in an introductory room, and measuring their behavior and reaction to the artwork on display in that room.

After weighting all these options, I decided to model the system with minimal base requirements, so that options could be easily added later, as a refinement to the system, by taking advantage of the modeling flexibility of Bayesian networks I explained in the previous section. These are the working hypothesis used to guide modeling the Bayesian network to estimate the visitor's type:

- 1. The information available to the system, modeled by the observed nodes of the network is location (which object the visitor is close by) and how long the visitor stays at each location (duration).
- 2. The priors on the three busy/greedy/selective type start equally for the three types. The visitor does not declare belonging to a user type at the entrance so as not to bias the priors on the visitor type to any of the three types. Some visitors might in fact feel bothered being asked or being committed to a type at start. When the museum wearable will be installed in a museum, it will be possible at the end of each week (for example) to count the number of busy/greedy/selective types that have been through the galleries and use this information as a prior on the visitor types for the following week. It is therefore a reasonable assumption for the first prototype presented in this document to start with equal priors on the three types.
- 3. There is no introductory space that allows the system to know the visitor better. While it would improve accuracy of type estimation at start, it would also impose a definite constraint on the curator and the exhibit designer, as such an introductory space would have to be found inside the museum galleries and set up. If such space became available it would allow the museum wearable to estimate a visitor type ahead of time during the visit. Without this prior estimation of the visitor type, the Bayesian network will make informed guesses about the visitor's interest along the exhibit. In time, as the visitor sees more objects the probabilistic nature of the system makes the guesses become more accurate.
- 4. Another assumption being made, is that each visitor belongs to one type which the system estimates during the visit. Because of the probabilistic nature of the system, this assumption does allow the system to change its mind about the visitor type if the visitor's behavior changes during the visit. What this assumption means

is that I have chosen the visitor type as a static node rather than a dynamic node. It can change but it is not a variable that evolves in time. This is to a certain extent a subjective choice of the modeler (myself), guided by commonsense and previous modeling experience. What it means is that even if the visitor starts with a busy behavior and later behaves selectively, the system can still account for that while considering the visitor as a unit and not as an entity whose value we sample in time. This assumption will be illustrated by some of the following Bayesian model examples.

- 5. To initially simplify the task at hand I have selected a subset of twelve representative objects at the MIT Museum's Robots and Beyond Exhibit, as shown in figure 42.

Once the basic modeling assumptions have been made various choices are still available when designing or modeling a Bayesian network for the exhibit. A variety of modeling techniques is described in [Jensen, 1996]. It is clear that the model needs to be some kind of dynamic Bayesian network, as estimating the visitor's type is a process that happens in time during the visit. Sampling times are given in this case by the presence of the visitor at each location, and therefore we need to model a process with twelve time slices, as twelve are the selected objects for the exhibit. In addition to this, the actual placement of these objects on the museum floor dictates some constraints on the model architecture. Therefore the geography of the objects in the exhibit needs to be reflected in the topology of the modeling network, as shown in figure 43. I will now describe the modeling steps and choices which have led to the Bayesian network which estimates the visitor's type from the location and stop duration information obtained from the infrared location sensor.

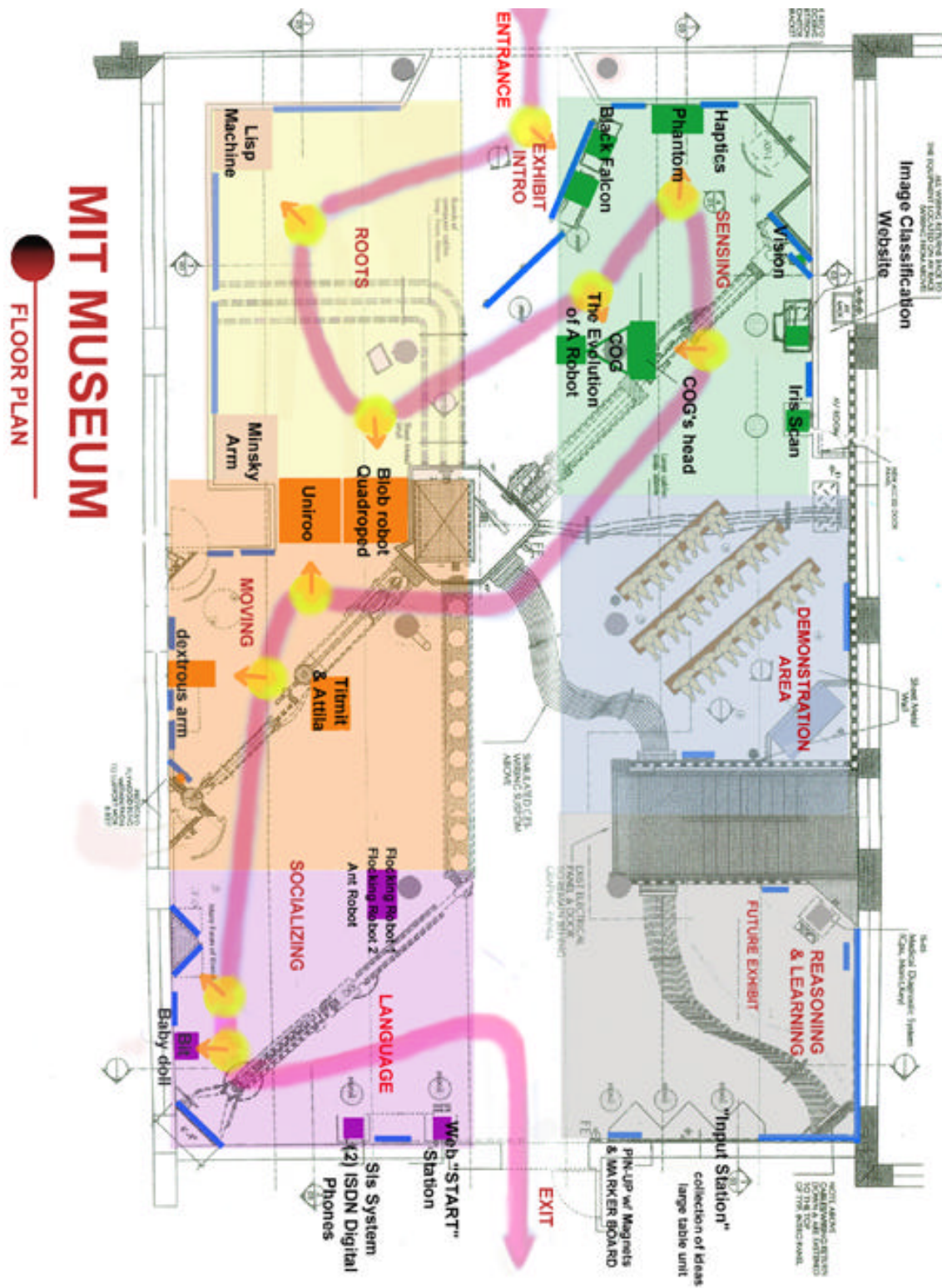


Figure 42. Selected objects at the MIT Museum's Robots and Beyond Exhibit

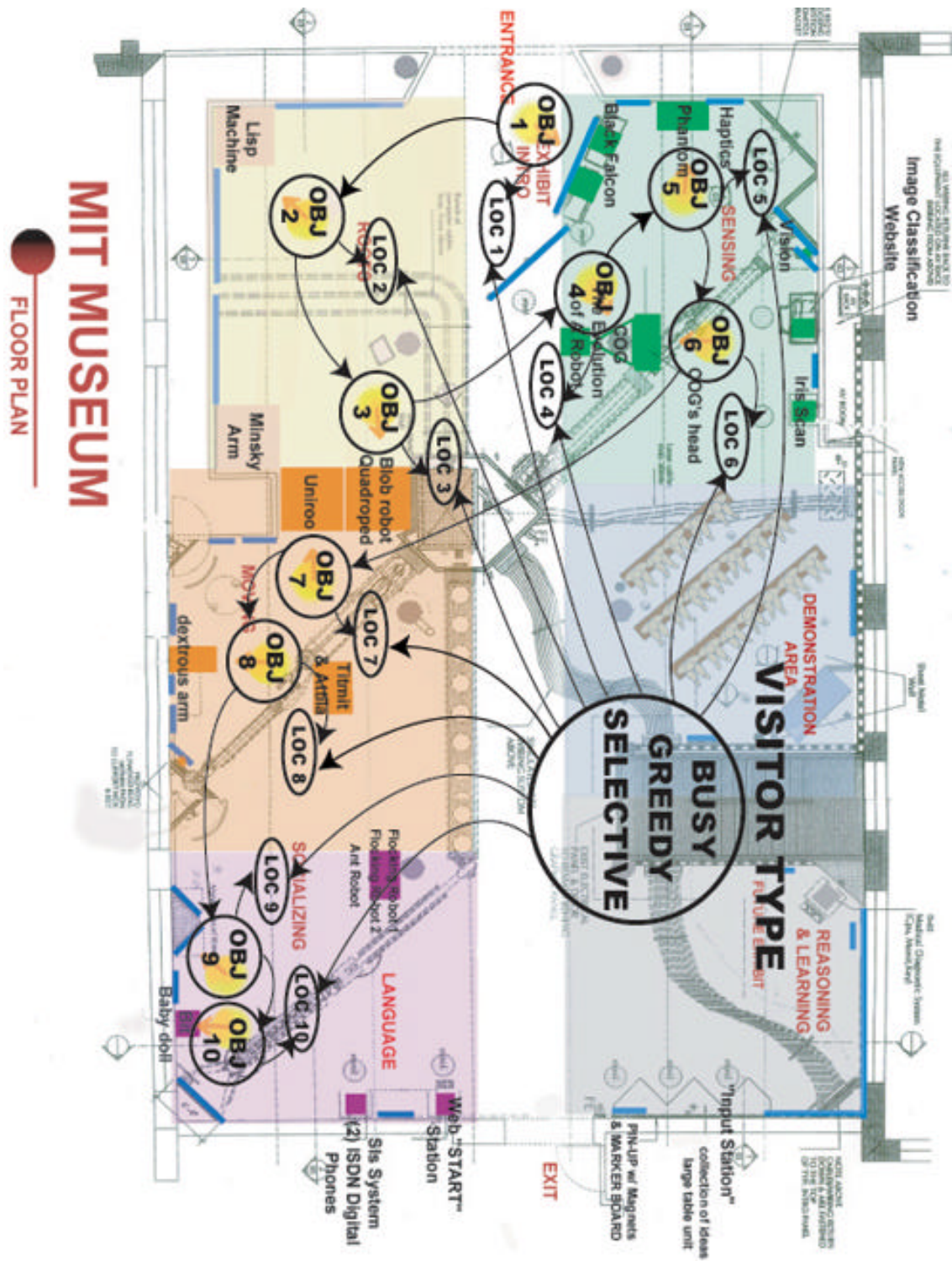


Figure 43. The geography of the exhibit needs to be reflected into the topology of the network

The choices available in modeling this problem are:

- the topology of the network
- which states per node
- whether to have a continuous or discrete node for the visitor's stop durations
- whether to have one unique node or multiple nodes to describe the stop durations.

These choices will affect the possible networks which model the problem. The first, somewhat easy, decision is to have discrete states for the visitor's stop durations. For simplicity we can choose three simple significant labels: 'skip', 'short', and 'long'. It is of course possible later on to extend the model to consider a higher resolution of discrete stop durations, such as one which includes: 'skip', 'short', 'medium', 'long', and 'very long'. I describe in section 7.2. a classification technique that allows the system designer to choose which range of values define these three classes. In brief, the ranges are specific to each museum exhibit and are derived from tracking data obtained by observing and measuring visitor's stop length and type at the museum, as described in section 3.5.

The expert's assumptions about the percentage of skip/short/long stop durations that a busy/greedy/selective will do are given in table 4. This table assigns numeric values to the qualitative description of these three types given at the beginning of this section. Section 7.1. shows how these a priori values can later be revised, and tuned to a specific exhibit, using the visitor tracking data described in section 3.5.

Conditional probability table for the visitor node			
	skip	short	long
Busy	0.2	0.7	0.1
Greedy	0.1	0.1	0.8
Selective	0.4	0.2	0.4

Table 4. Conditional Probability Table for the visitor node

I will now describe four possible networks and explain the selection criteria which has led to the final choice.

#### 4.2.1. Visitor Type model 1

The first model is a naive first tentative which weights each possibility for a busy/greedy/selective type to make a skip/short/long stop duration. Basically for each 'skip', 'short', 'long' event, modeled as separate nodes, it defines the probability that it is caused by a busy/greedy/selective type. To simplify the assignment of the conditional probability tables for the binary (true/false) of the skip/short/long nodes it is convenient to use a noisy-or modeling for each of these nodes. When a variable B has several parents, one must specify  $p(B | A^*)$  for each configuration  $A^*$  of the parents. In this case for all binary nodes, this would imply having to assign:  $p(\text{duration} | \text{busy, greedy, selective})$  for all combinations of busy/greedy/selective. Specifying such a configuration



may be too specific for any expert. A simplifying method is provided by the noisy-or modeling technique [Jensen, 1996]. It is based on the assumption that an event happens unless something prevents it to happen.

If  $A_1, \dots, A_n$  are binary variables listing all the causes of the binary variable  $B$ , each event  $A_i = y$  causes  $B = y$  unless an inhibitor prevents it, and the probability for that is  $q_i$ . This is expressed by:  $p(B = n | A_i = y) = q_i$ . Then:  $p(B = n | A_1, A_2, \dots, A_n) = \prod_{j \in Y} q_j$  where  $Y$  is the set of indices for variables in the state  $y$ . For example:  $p(B = y | A_1 = y, A_2 = y, A_3 = A_n = n) = 1 - q_1 q_2$  [figure 44].

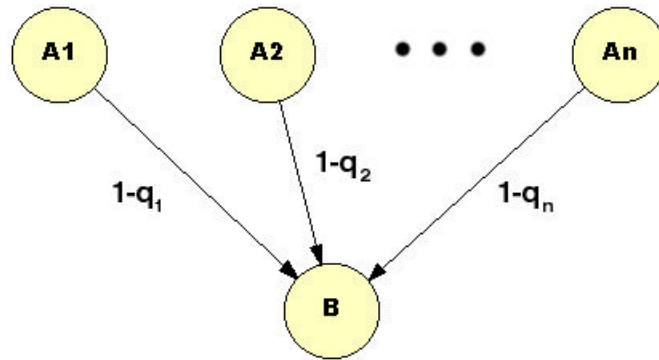


Figure 44. Noisy-or modeling

By using the noisy-or modeling technique the number of probabilities to assign grows only linearly with the numbers of parents.

For the museum visitor estimation problem the network is shown in figure 45, for one time slice and the conditional probability tables obtained by noisy-or modeling is shown in table 5. The priors  $p(\text{busy}=\text{yes})$ ,  $p(\text{greedy}=\text{yes})$  and  $p(\text{selective}=\text{yes})$  are all 0.3333, that is equal for all.

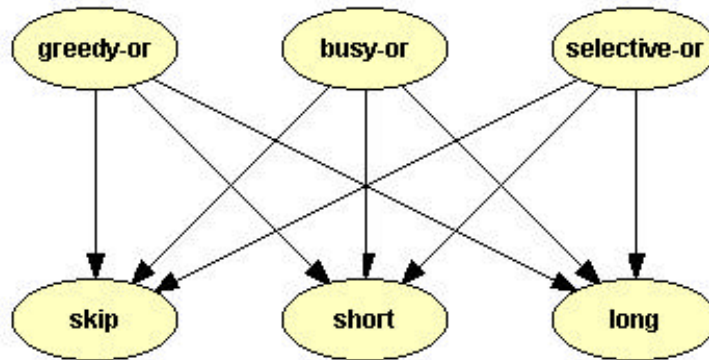


Figure 45. Topology of model 1, one time slice.

<b>p(SK SBG)</b>		
<b>sk=yes</b>	<b>sk=no</b>	<b>% SK=skip</b>
0.67	0.33	% S=selective B=busy G=greedy
0.63	0.37	% S=selective B=busy G=not greedy
0.59	0.41	% S=selective B=not busy G=greedy
0.54	0.46	% S=selective B=not busy G=not greedy
0.45	0.55	% S=not selective B=busy G=greedy
0.39	0.61	% S=not selective B=busy G=not greedy
0.30	0.69	% S=not selective B=not busy G=greedy
0.23	0.77	% S=not selective B=not busy G=not greedy
<b>p(SH SBG)</b>		
<b>sh=yes</b>	<b>sh=no</b>	<b>% SH=short</b>
0.86	0.14	% S=selective B=busy G=greedy
0.84	0.16	% S=selective B=busy G=not greedy
0.52	0.48	% S=selective B=not busy G=greedy
0.47	0.53	% S=selective B=not busy G=not greedy
0.82	0.18	% S=not selective B=busy G=greedy
0.80	0.20	% S=not selective B=busy G=not greedy
0.40	0.60	% S=not selective B=not busy G=greedy
0.33	0.67	% S=not selective B=not busy G=not greedy
<b>p(LO SBG)</b>		
<b>lo=yes</b>	<b>lo=no</b>	<b>% LO=long</b>
0.94	0.06	% S=selective B=busy G=greedy
0.70	0.30	% S=selective B=busy G=not greedy
0.93	0.07	% S=selective B=not busy G=greedy
0.66	0.34	% S=selective B=not busy G=not greedy
0.90	0.10	% S=not selective B=busy G=greedy
0.49	0.51	% S=not selective B=busy G=not greedy
0.89	0.11	% S=not selective B=not busy G=greedy
0.43	0.57	% S=not selective B=not busy G=not greedy

Table 5. Conditional Probability tables for the SKIP, SHORT, and LONG nodes resulting from the noisy-or modeling

The transition probabilities between one time slice and the text are given by table 6. The transition probabilities are the same for models 1, 2, and 3.

	<b>greedy</b>	<b>not greedy</b>
<b>greedy</b>	0.6	0.4
<b>not greedy</b>	0.4	0.6
	<b>busy</b>	<b>not busy</b>
<b>busy</b>	0.6	0.4
<b>not busy</b>	0.4	0.6
	<b>selective</b>	<b>not selective</b>
<b>selective</b>	0.6	0.4
<b>not selective</b>	0.4	0.6

Table 6. Transition Probability tables for the GREEDY, BUSY, and SELECTIVE nodes

The meaning of the transition probability table is quite important for the system. The values along the diagonal describe how much past history matters in determining if a visitor belongs to a type. For examples if the values along the diagonal are high (0.9) and the visitor switches behavior from let say greedy to selective, the system will have some inertia in identifying the new type. It will take the system a few time steps to catch up with the new visitor behavior. This makes sense if there are many objects in the exhibit. If instead there are only a few objects in the exhibit, such as MIT's Robots and Beyond targeted exhibit, it is desirable to have a system which adapts fast to the visitor's behavior and does not tend to "stick to the first impression" as before. In this case it is preferable to have lower values along the diagonal (0.6) as shown in table 6.

To test the model, I introduced evidence on the duration nodes, thereby simulating its functioning during the museum visit. I have included results below, limited to two time slices, for the limited space available on paper. The reader can verify that the system gives plausible estimates of the visitor type, based on the evidence introduced in the system. The posterior probabilities in this and the subsequent models are calculated using Hugin, which implements the Distribute Evidence and Collect Evidence message passing algorithms on the junction tree.

Summary of results:

**Test case 1.** The visitor spends a short time both with the first and second object → the network gives the highest probability to the *busy* type (0.6705)

**Test case 2.** The visitor spends a long time both with the first and second object → the network gives the highest probability to the *greedy* type (0.5943)

**Test case 3.** The visitor spends a long time with the first object and skips the second object → the network gives the highest probability to the *selective* type (0.6016)

The importance of the transition probability table can be verified on the simple test case 3. Figures 46-51 show results obtained with a transition probability with high inertia (high values along the diagonal = 0.9) and with low inertia (diagonal probabilities = 0.6). I introduce evidence separately for time slice 1 and 2. When I introduce evidence for a long stop at object 1, for time slice 1, both systems select the greedy type as the highest probability estimate for the visitor type, which is consistent with the type definition. However when I introduce evidence for skip at time step 2, the systems make slightly different inferences. The system with low inertia stays with the type estimate made at time step 1, and basically says "at time step 1 I had a greedy visitor, but now after a skip I think that the highest probability is for a selective type". The system with high inertia instead changes its mind about what had happened at step 1, and says something like "well given that now at time step 2 I observe a selective visitor, I must actually have been wrong at time step one, and I am going to correct my estimate for step 1 from greedy to selective". This is an example of "explaining away" probabilistic reasoning, which causes the probability of greedy at step 1 to lower as it is explained by the selective type at step 2. The reader has been introduced to explaining away in the examples provided in section 4.1. Ultimately there is really no right or wrong way of setting the transition probability table. It simply depends on the problem at hand, and the choices that the modeler and the curator find appropriate on the basis of the public they address and the message they wish to convey.



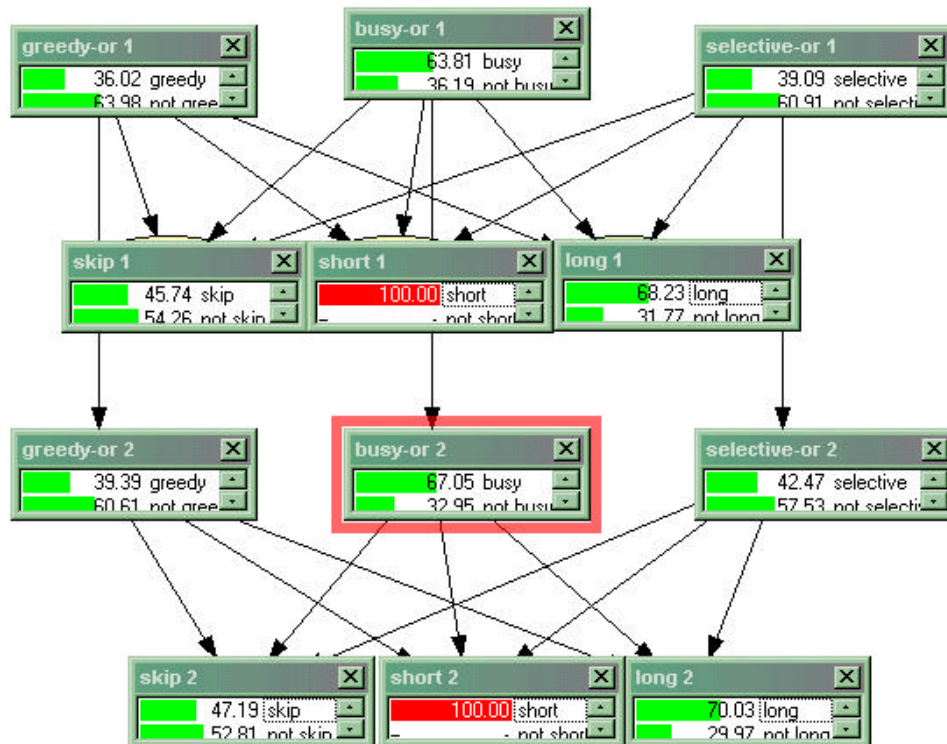


Figure 46. Posterior probabilities after evidence is introduced and inference is performed (Duration 1=short and Duration 2=short)

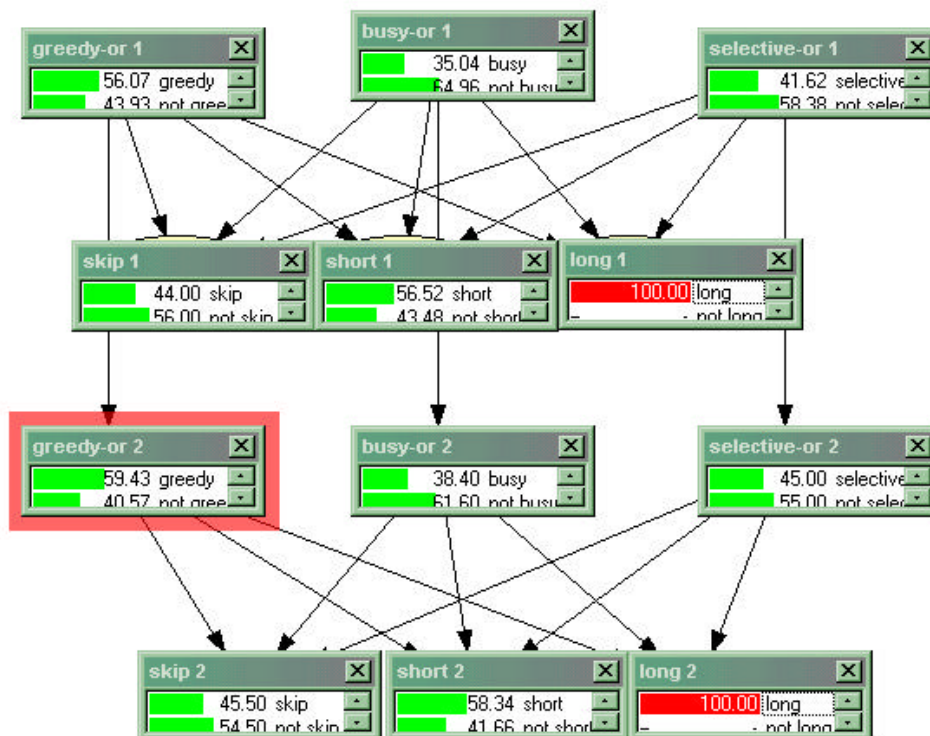


Figure 47. Posterior probabilities after evidence is introduced and inference is performed (Duration 1=long and Duration 2=long)

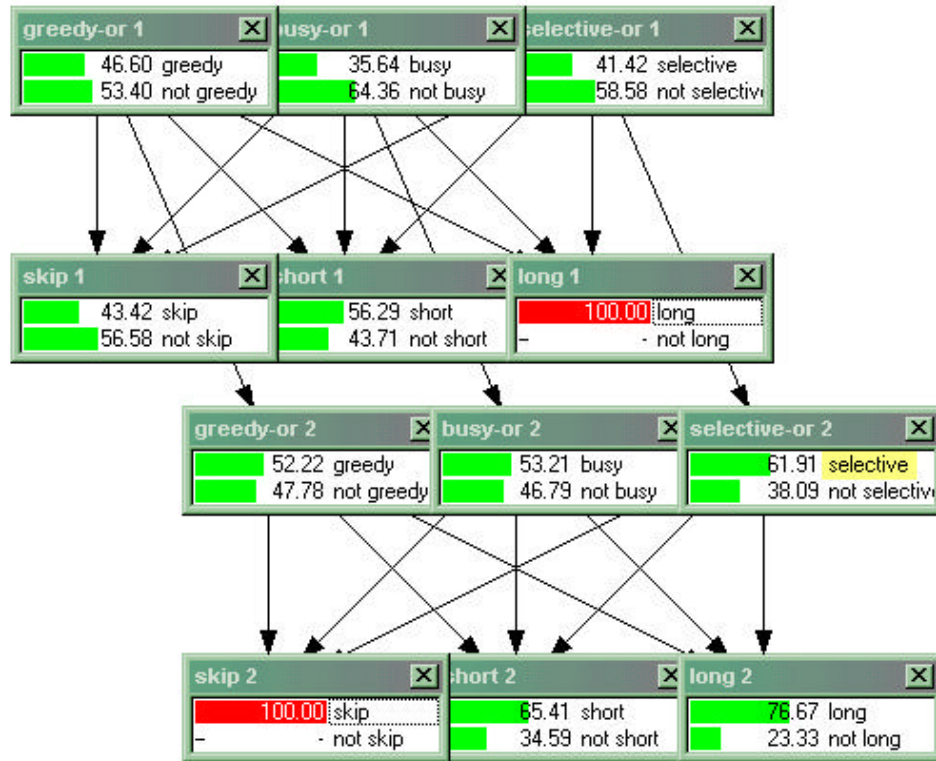


Figure 48. Posterior probabilities after evidence is introduced and inference is performed, for model 1, (Duration 1=long and Duration 2=short) – system with \*low\* inertia.

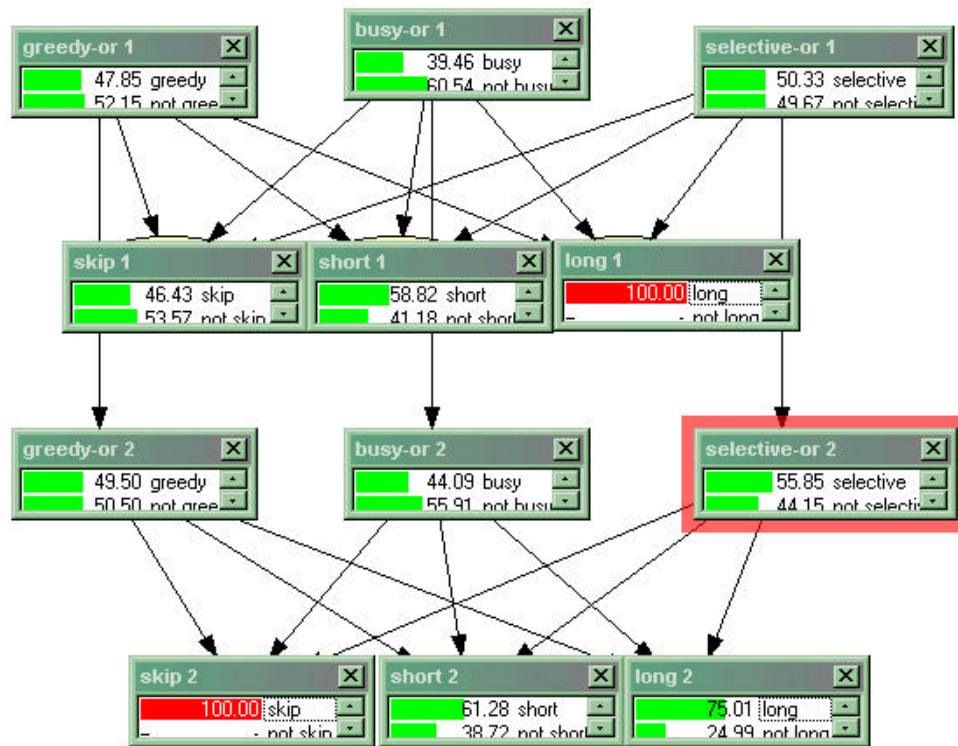


Figure 49. Posterior probabilities after evidence is introduced and inference is performed, for model 1, (Duration 1=long and Duration 2=short) – system with \*high\* inertia.

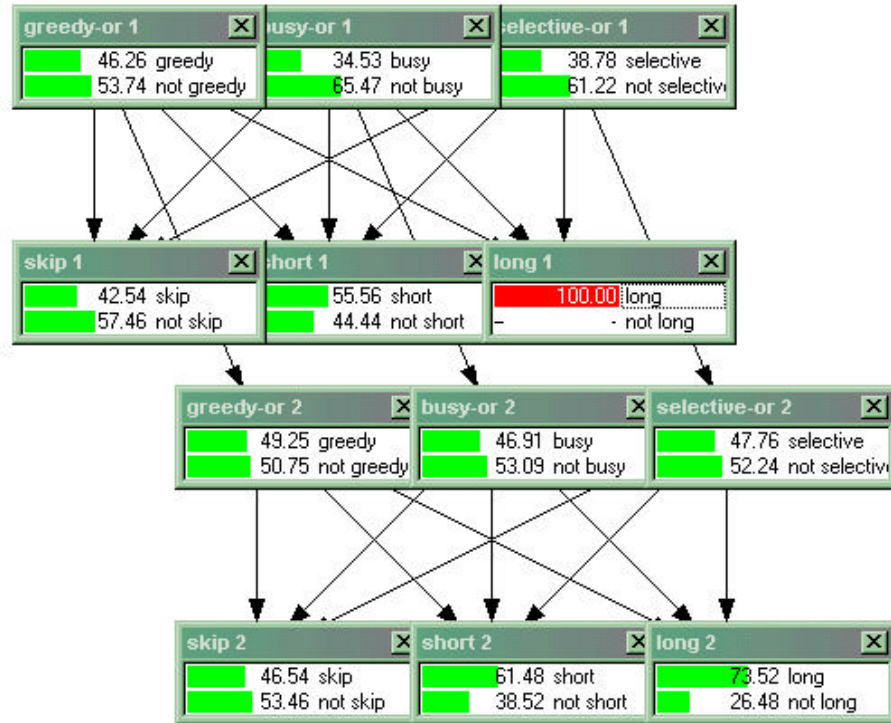


Figure 50. Posterior probabilities after evidence is introduced and inference is performed, for model 1, (Duration 1=long) – system with "low" inertia.

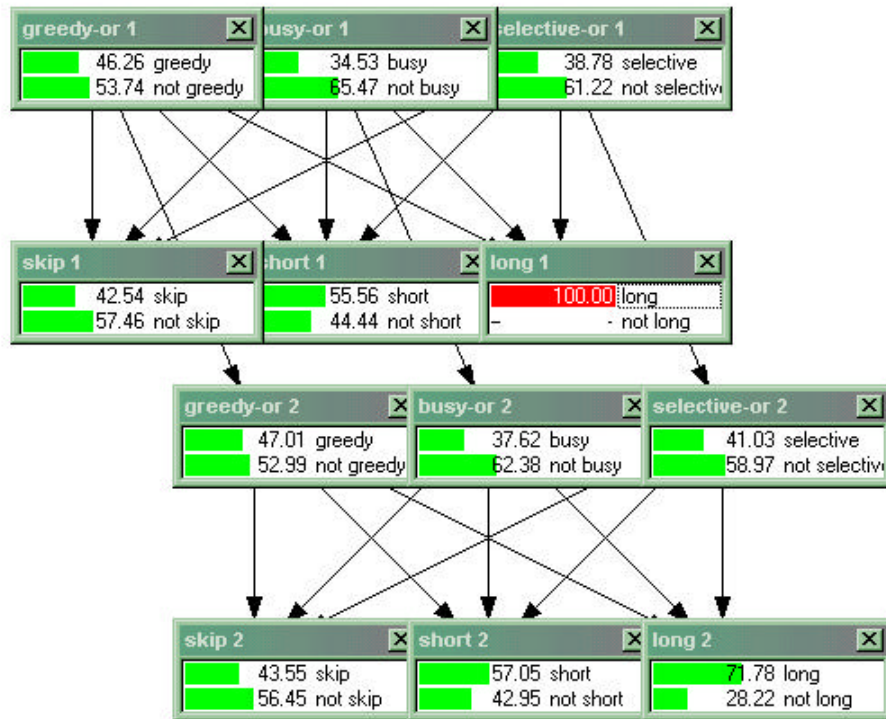


Figure 51. Posterior probabilities after evidence is introduced and inference is performed, for model 1, (Duration 1=long) – system with "low" inertia.

### 4.2.2. Visitor Type model 2

An alternative to the previous model, can be found by thinking what the different visitor types have in common. For example both the busy and selective type tend to skip, the selective or greedy do not make short stops, whereas the busy type makes many short stops. The model shown in figure 52 extracts all similarities and differences between types and encodes them in the topology and parameters of the model. It is

of importance to notice that the (type or type) nodes do not necessarily have connections to all skip/short/long nodes. The whole point of this model is that it is somewhat of a summary of what common types might do, and therefore we do not need edges to all leaf nodes. This allows us to simplify the conditional probability tables for the leaf nodes with less incoming edges [tables 7,8,9]. As before, to simplify the assignment of the conditional probability tables for the duration nodes, I used noisy-or to model the leaf nodes. All the (type or type) nodes are logical XOR nodes. The priors  $p(\text{busy}=\text{yes})$ ,  $p(\text{greedy}=\text{yes})$  and  $p(\text{selective}=\text{yes})$  are all 0.3333, that is equal for all. The test cases illustrates in figures 53-55 are the same as in the previous model and all lead to plausible results.

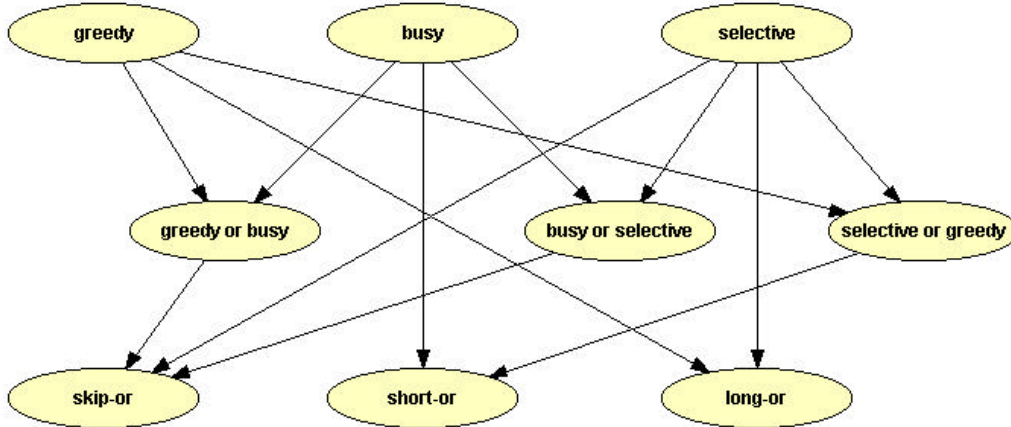


Figure 52. Topology of model 2, one time slice.

$p(\text{SK}   \text{S BoS GoB})$		
sk=yes	sk=no	
0.73	0.27	% S=selective BoS=bars GoB=gorb
0.68	0.32	% S=selective BoS=bars GoB=s(not)(gorb)
0.60	0.39	% S=selective BoS=g not(bars) GoB=gorb
0.54	0.46	% S=selective BoS=g not(bars) GoB=s(not)(gorb)
0.54	0.46	% S=notselective BoS=bars GoB=gorb
0.46	0.54	% S=notselective BoS=bars GoB=s(not)(gorb)
0.35	0.65	% S=notselective BoS=g not(bars) GoB=gorb
0.23	0.77	% S=notselective BoS=g not(bars) GoB=s(not)(gorb)

Table 7. Conditional Probability tables for the SKIP nodes, uses noisy-or



p(SH   B SoG)		
sh=yes	sh=no	
0.83	0.17	% B=busy SoG=s.org
0.80	0.20	% B=busy SoG=b not(s.org)
0.43	0.57	% B=not busy SoG=s.org
0.33	0.67	% B=not busy SoG=b not(s.org)

Table 8. Conditional Probability tables for the SHORT nodes, uses noisy-or

p(LO   SG)		
lo=yes	lo=no	
0.93	0.07	% S=selective G=greedy
0.66	0.34	% S=selective G=not greedy
0.89	0.11	% S=not selective G=greedy
0.43	0.57	% S=not selective G=not greedy

Table 9. Conditional Probability tables for the LONG nodes, uses noisy-or

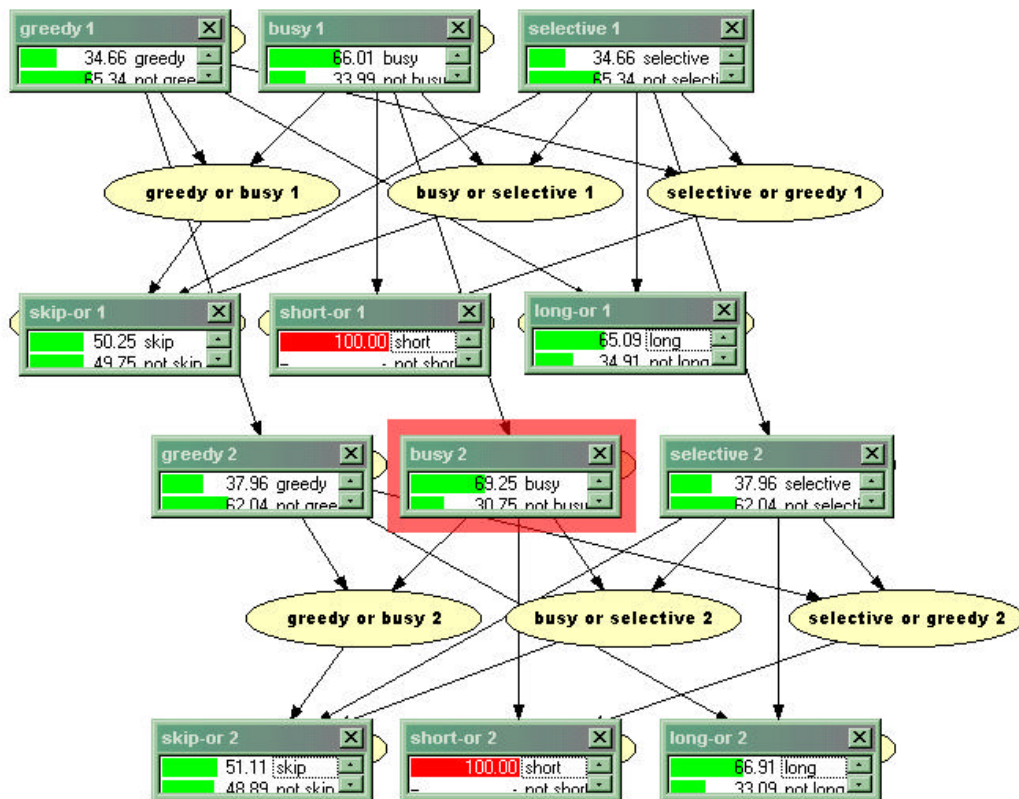


Figure 53. Posterior probabilities after evidence is introduced and inference is performed, for model 2.

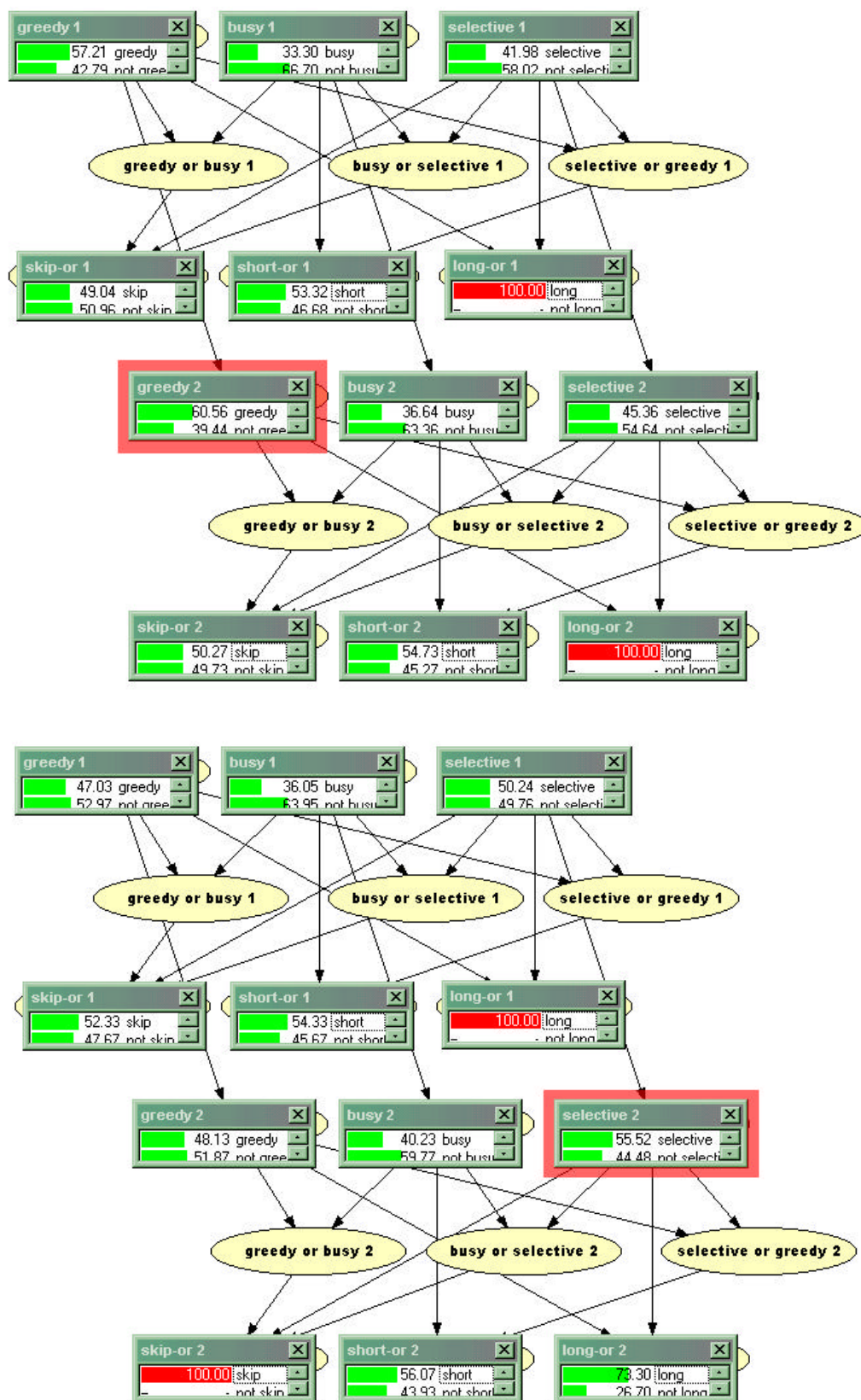


Figure 54, 55. Posterior probabilities after evidence is introduced and inference is performed, for model 2.

In summary, for model 2:

**Test case 1.** The visitor spends a short time both with the first and second object → the network gives the highest probability to the *busy* type (0.6925)

**Test case 2.** The visitor spends a long time both with the first and second object → the network gives the highest probability to the *greedy* type (0.6056)

**Test case 3.** The visitor spends a long time with the first object and skips the second object → the network gives the highest probability to the *selective* type (0.5552)

### 4.2.3. Visitor Type model 3

Model 3 performs an inversion of the causal arrows with respect to the previous models. It is equally possible to imagine that a greedy visitor generates a long stop, as well as to think that a long stop “generates” a greedy visitor. For the sake of exploration I also created a model with inverted causal arrow and obtained once again plausible results [figure 56]. As before, in figures 57-59 the reader will find probability update calculation for a simple two time-slices test case. The conditional probability tables for greedy/busy/selective are also given in tables 10,11,12.

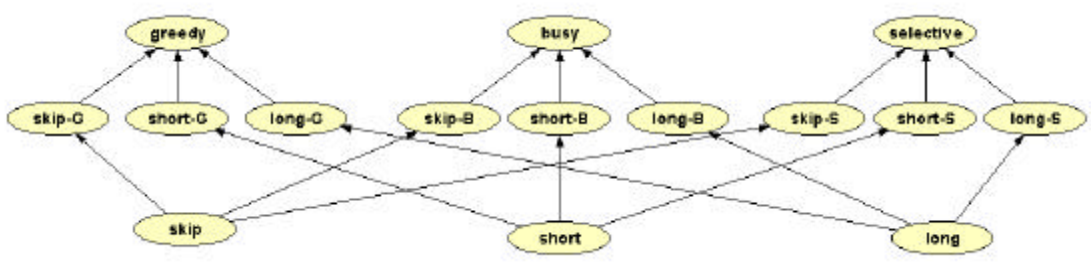


Figure 56. Topology of model 3, one time slice.

$p(B   SKB LOB SHB)$		
$b=yes$	$B=no$	
0.02	0.98	% $SKB=skip-B LOB=long-B SHB=short-B$
0.01	0.99	% $SKB=skip-B LOB=long-B SHB=not short-B$
0.75	0.25	% $SKB=skip-B LOB=not long-B SHB=short-B$
0.58	0.42	% $SKB=skip-B LOB=not long-B SHB=not short-B$
0.12	0.88	% $SKB=not skip-B LOB=long-B SHB=short-B$
0.06	0.94	% $SKB=not skip-B LOB=long-B SHB=not short-B$
0.94	0.06	% $SKB=not skip-B LOB=not long-B SHB=short-B$
0.89	0.11	% $SKB=not skip-B LOB=not long-B SHB=not short-B$

Table 10. Conditional Probability table for the busy node.

<b>p(G   SKG LOG SHG)</b>		
<b>g=yes</b>	<b>G=no</b>	
0.01	0.99	% SKG=skip-G LOG=long-G SHG=short-G
0.34	0.66	% SKG=skip-G LOG=long-G SHG=not short-G
0.01	0.99	% SKG=skip-G LOG=not long-G SHG=short-G
0.22	0.78	% SKG=skip-G LOG=not long-G SHG=not short-G
0.14	0.86	% SKG=not skip-G LOG=long-G SHG=short-G
0.89	0.11	% SKG=not skip-G LOG=long-G SHG=not short-G
0.09	0.91	% SKG=not skip-G LOG=not long-G SHG=short-G
0.81	0.18	% SKG=not skip-G LOG=not long-G SHG=not short-G

Table 11. Conditional Probability table for the greedy node.

<b>p(S   LOS SKS SHS)</b>		
<b>s=yes</b>	<b>s=no</b>	
0.05	0.95	% LOS=long-S SKS=skip-S SHS=short-S
0.47	0.53	% LOS=long-S SKS=skip-S SHS=not short-S
0.06	0.94	% LOS=long-S SKS=not skip-S SHS=short-S
0.52	0.48	% LOS=long-S SKS=not skip-S SHS=not short-S
0.39	0.61	% LOS=not long-S SKS=skip-S SHS=short-S
0.92	0.08	% LOS=not long-S SKS=skip-S SHS=not short-S
0.44	0.56	% LOS=not long-S SKS=not skip-S SHS=short-S
0.93	0.07	% LOS=not long-S SKS=not skip-S SHS=not short-S

Table 12. Conditional Probability table for the selective node.

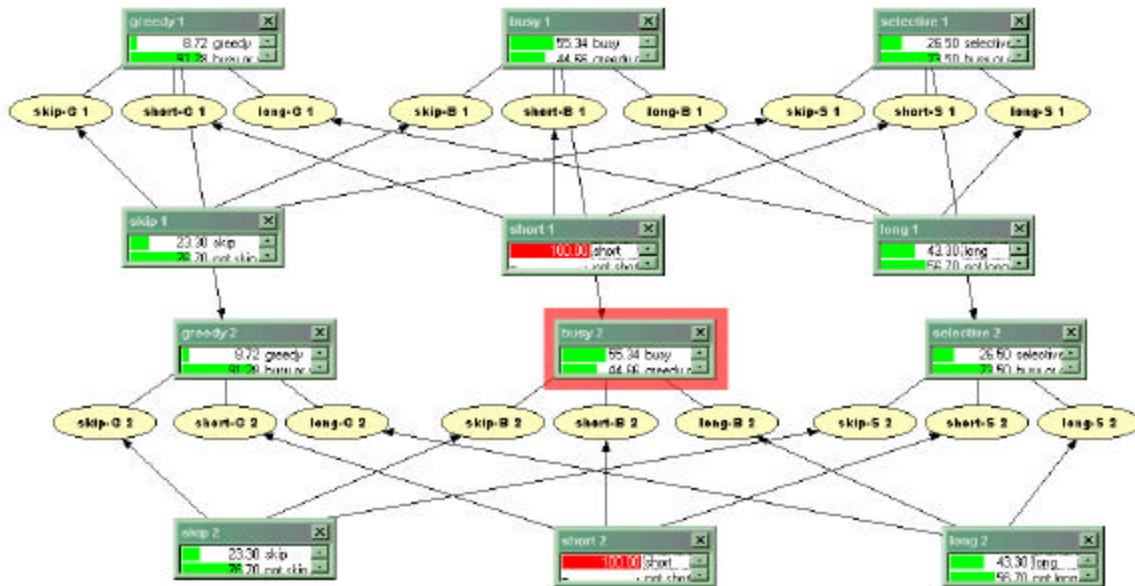
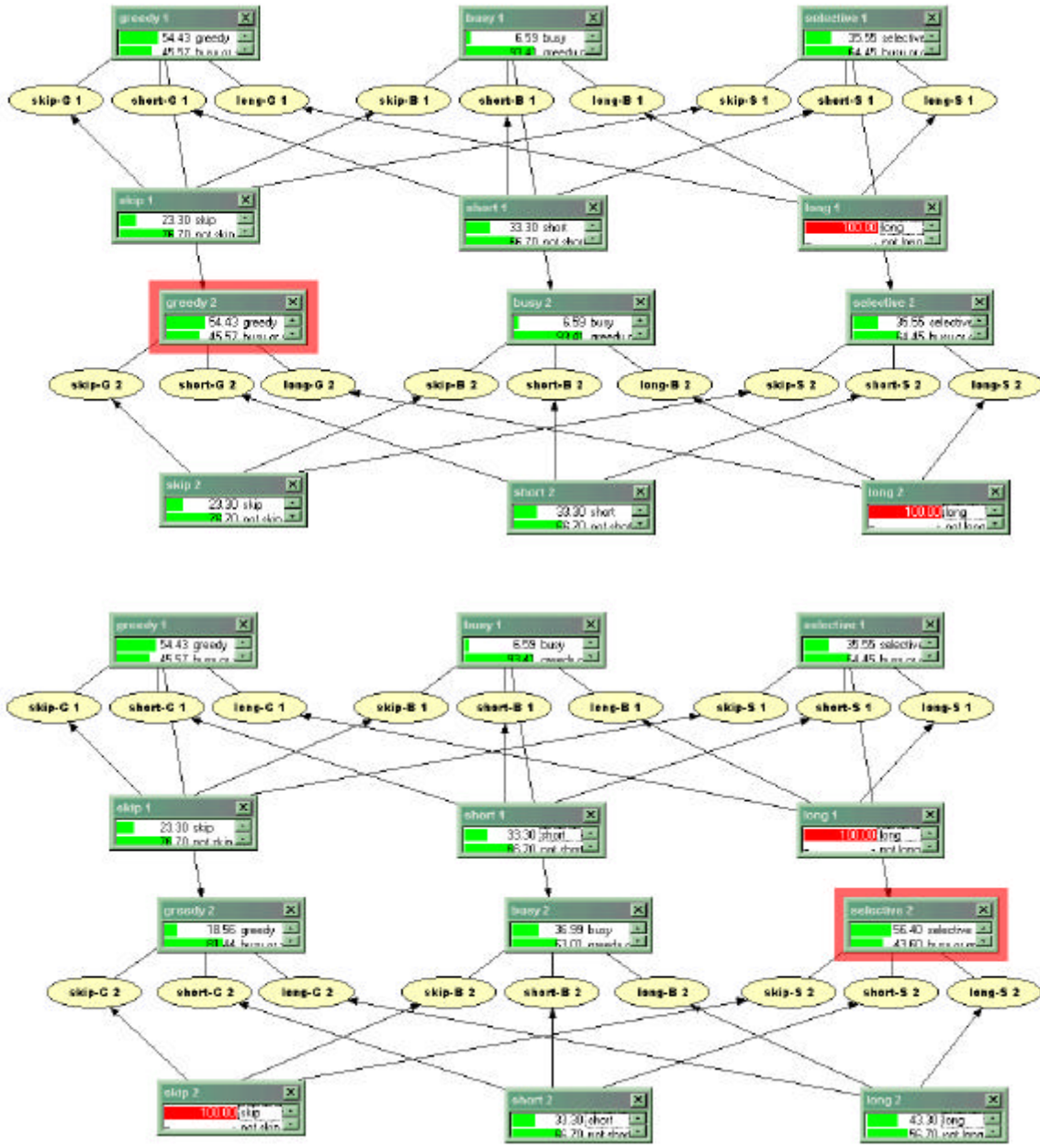


Figure 57. Posterior probabilities after evidence is introduced and inference is performed, for model 3.





Figures 58, 59. Posterior probabilities after evidence is introduced and inference is performed, for model 3.

The summary of results for model 3 is:

**Test case 1.** The visitor spends a short time both with the first and second object → the network gives the highest probability to the *busy* type (0.5534)

**Test case 2.** The visitor spends a long time both with the first and second object → the network gives the highest probability to the *greedy* type (0.5443)

**Test case 3.** The visitor spends a long time with the first object and skips the second object → the network gives the highest probability to the *selective* type (0.5640)

#### 4.2.4. Visitor Type models 4a and 4b

A final possibility is to consider a model which has the topology of an HMM, such as the one shown in figure 60. However such a model would require having a dynamic visitor node, which is a choice I discarded in the above discussion on the working hypothesis for this model. A similar topology, but with a static visitor node, can be obtained by introducing an object node which actually encodes information about the object being observed by the visitor. For example, some objects can be very interesting or less interesting, according either to the opinion of the curator, or the public's preferences, and this information can be encoded in the Bayesian network. Moreover, in a situation in which the museum wearable was available to the public, it would be possible, theoretically, at the end of each week, to take the posterior probabilities for all objects and all visitors, and reintroduce them as priors for the Bayesian network used in the wearables the following week. This would allow the system to account for the evolving public's preferences and types. While installation of the museum wearable in the museum is not addressed in the research described by this document, it is important to notice the possibility to easily update the priors of the model from the posterior probabilities of the previous day or week to obtain a more accurate estimate of the variables of interest.

To obtain a better network topology, another possibility is to group together the binary skip/short/long nodes, which only have true/false states, as in the examples above, into one unique location node which represents how much time the visitor spends at that location, i.e. with a ternary skip/short/long state [figure 61]. This is well representative of the problem as truly an object skip excludes a short stop or a long stop and viceversa. If skip/short/long are states of the same node the sum of their total probability needs to be one, which mathematically translates the previous statement.

I initially implemented the latter model [figure 62], for simplicity, with a binary object state with visited/not\_visited states. The reason is that for the training of the model I later performed, and which is discussed in section 7.1., no priors were available on whether the objects were more interesting than neutral or boring, and I would have had to learn separate conditional probabilities for the neutral/interesting/boring cases, for which no training data was available.

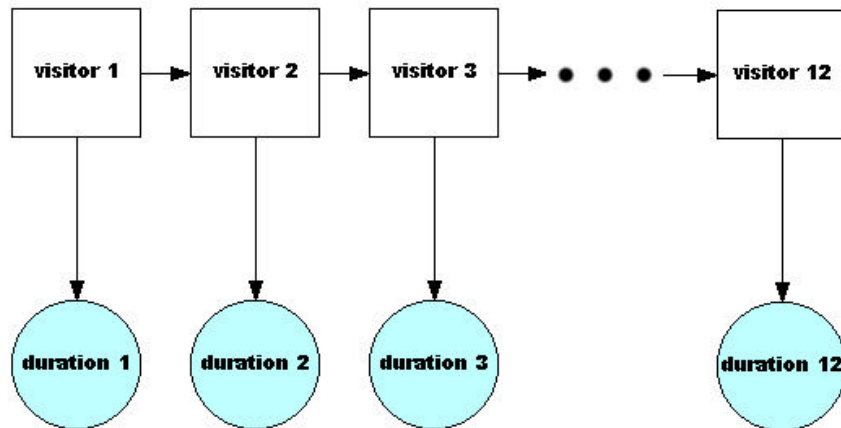


Figure 60. HMM modeling of the visitor type estimation problem.

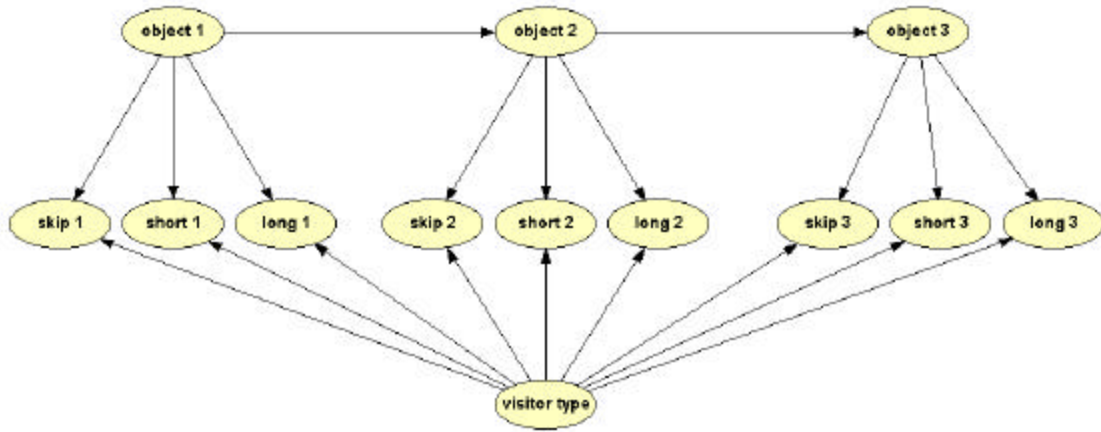


Figure 61. Model 4a, which introduces object nodes. Only 3 time slices out of 12 are shown in figure.

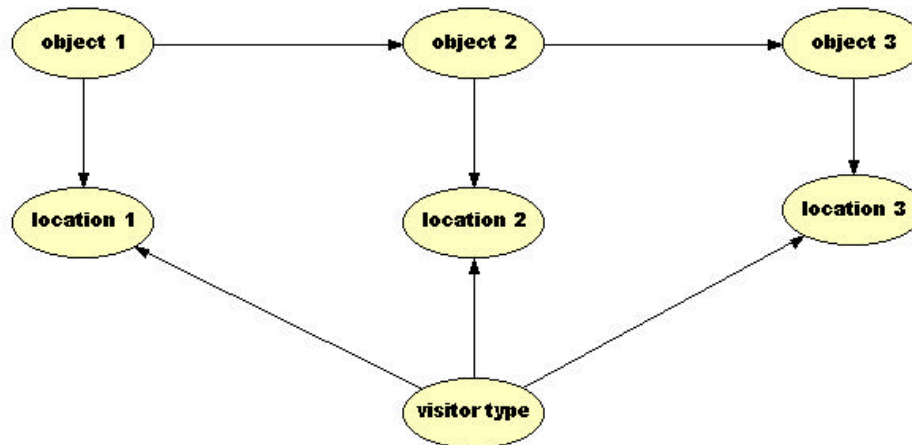


Figure 62. Final selection: model 4b, with object nodes and ternary skip/short/long location nodes, 3 time slices.

After having taken tracking data on the twelve selected objects, as described in section 3.5, it is possible to analyze this data and infer priors on whether the objects are neutral/interesting/boring. Then we need to give different conditional probabilities for the interesting and boring cases, such that if a busy type spends quite a long type with an object, that is more an indication of that object being interesting than the busy type becoming more greedy in their visit.

Table 13 shows the total amount of time, in seconds, that the fifty tracked visitors have spent at the twelve targeted objects at MIT's Robots and Beyond exhibit:

<b>Intro</b> <b>1</b>	<b>Lisp</b> <b>2</b>	<b>Minsky</b> <b>Arm 3</b>	<b>Robo</b> <b>Arm 4</b>	<b>Falcon</b> <b>5</b>	<b>Phantom</b> <b>6</b>	<b>CogsHead</b> <b>7</b>	<b>Quad</b> <b>8</b>	<b>Uniroo</b> <b>9</b>	<b>Dext</b> <b>Arm 10</b>	<b>Kismet</b> <b>11</b>	<b>Baby Doll</b> <b>12</b>
588	755	719	627	986	748	771	694	859	700	1025	548

Table 13. Total time spent by the 50 visitors at the 12 observed objects

A simple histogram plot [figure 63] shows which of these objects are interesting/neutral/boring, always in a probabilistic sense. The tallest bins are from objects 5,9,11, which are therefore the most interesting. The shortest bins are from objects 1,4,12 that are therefore boring. All remaining objects shall be considered neutral.

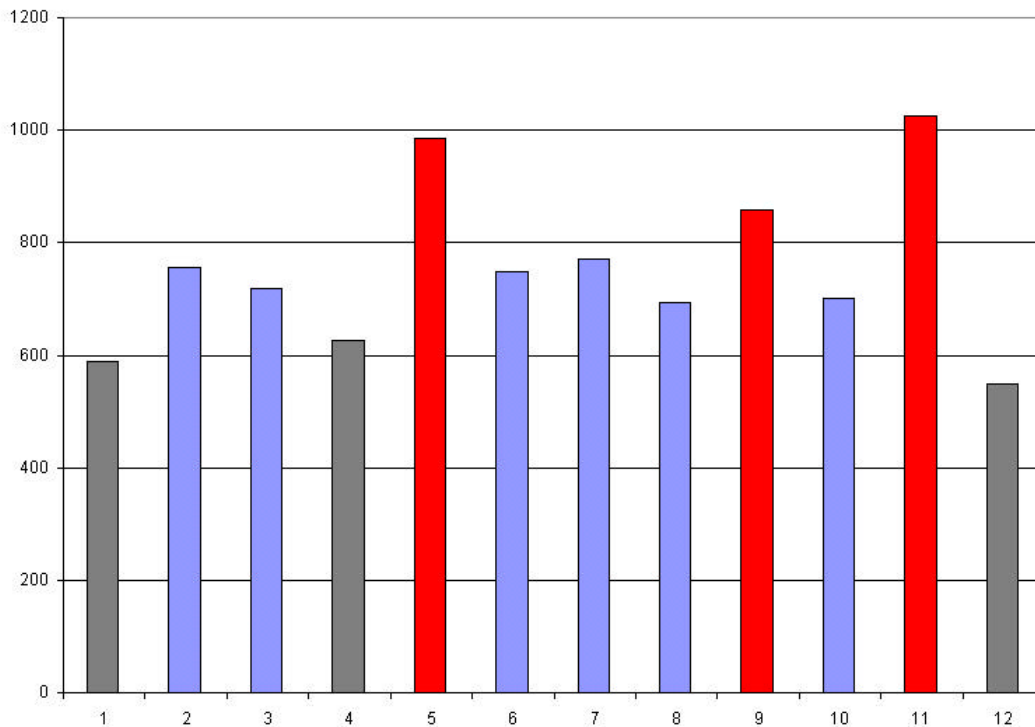


Figure 63. Histogram of overall time, in seconds, that visitors spend at the 12 target objects.

Based on the information above, a set of priors of the twelve objects is given by table 14:

	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	<b>11</b>	<b>12</b>
<b>neutral</b>	0.23	0.34	0.28	0.25	0.18	0.32	0.36	0.24	0.27	0.26	0.16	0.2
<b>interesting</b>	0.23	0.33	0.36	0.25	0.64	0.34	0.32	0.38	0.46	0.37	0.68	0.2
<b>boring</b>	0.54	0.33	0.36	0.5	0.18	0.34	0.32	0.38	0.27	0.37	0.16	0.6

Table 14. Priors for the 12 selected objects derived from the visitor tracking data.

The priors on the visitor states are busy=greedy=selective=0.333. The conditional probabilities  $p(\text{location} | \text{object}, \text{visitor})$  is shown in table 15:

	neutral			interesting			boring		
	busy	greedy	selective	busy	greedy	selective	busy	greedy	selective
skip	0.2	0.1	0.4	0.1	0.05	0.1	0.35	0.2	0.65
short	0.7	0.1	0.2	0.6	0.05	0.3	0.6	0.2	0.15
long	0.1	0.8	0.4	0.3	0.9	0.6	0.05	0.6	0.2

Table 15. Conditional Probability Tables for the Location Nodes.

The initial object and transition probabilities from one object to the next are given by table 16:

P(O1)		P(Oj   Oi)	neutral	interesting	boring
neutral	0.333	neutral	0.6	0.2	0.2
interesting	0.333	interesting	0.2	0.6	0.2
boring	0.333	boring	0.2	0.2	0.6

Table 16. Priors for the 12 selected objects and transition probabilities.

The transition probabilities are in this model the same for all twelve objects. However these could be also learned if more training data was available. For example visitors could have a tendency to skip certain objects, or visit them in a different order than what the exhibit designer has laid out. Alternatively, the transition probabilities would highlight groups of object and show that for example the “Sensors” section of the exhibit turns out to be more interesting than the “Movement” section. This information would be reflected in the transition tables. From the visitor tracking data gathered at the MIT Museum we observed people visiting objects one after the next in a linear sequence. This observation is mapped to the table above.

The test cases illustrates in figures 64-66 are the same as in the previous models and all lead to plausible results. The summary of results for model 4 is:

**Test case 1.** The visitor spends a short time both with the first and second object → the network gives the highest probability to the *busy* type (0.8592)

**Test case 2.** The visitor spends a long time both with the first and second object → the network gives the highest probability to the *greedy* type (0.7409)

**Test case 3.** The visitor spends a long time with the first object and skips the second object → the network gives the highest probability to the *selective* type (0.5470)

I have also included a test case for the HMM-like version of the model discussed above, with the same  $p(\text{object}|\text{visitor})$  as model 4 [figures 67-69], which led to the following results:

**Test case 1.** The visitor spends a short time both with the first and second object → the network gives the highest probability to the *busy* type (0.8640)

**Test case 2.** The visitor spends a long time both with the first and second object → the network gives the highest probability to the *greedy* type (0.75)

**Test case 3.** The visitor spends a long time with the first object and skips the second object → the network gives the highest probability to the *selective* type (0.5874)

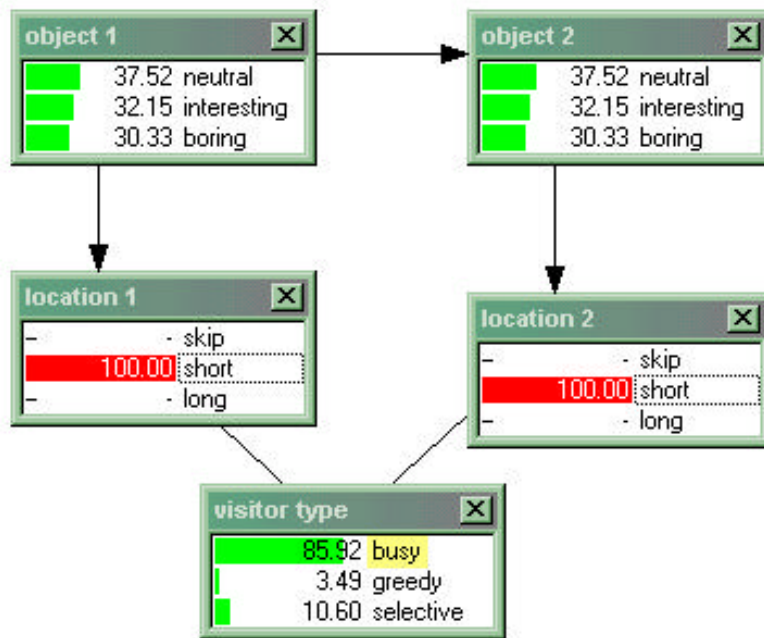


Figure 64. Posterior probabilities after evidence is introduced and inference is performed, for model 4: busy type.

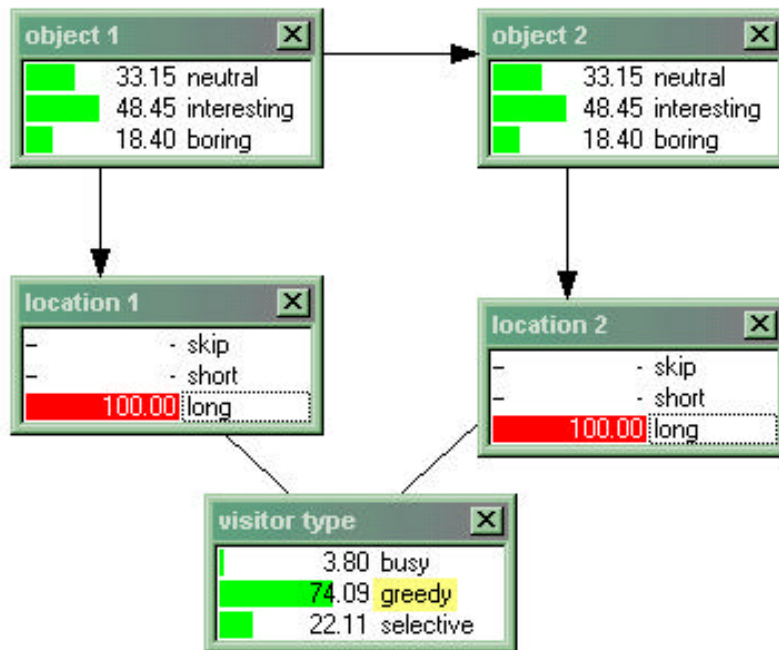


Figure 65. Posterior probabilities after evidence is introduced and inference is performed, for model 4: greedy type.

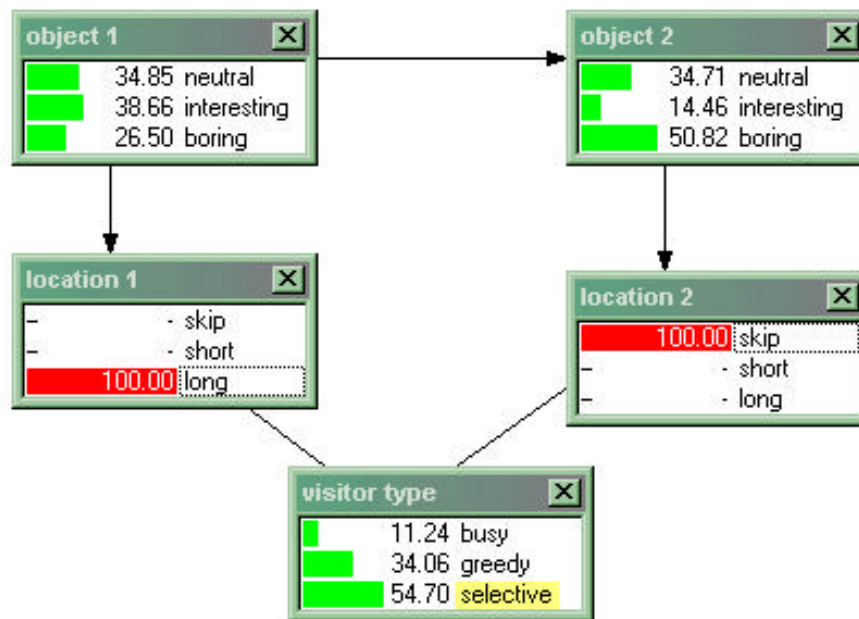


Figure 66. Posterior probabilities after evidence is introduced and inference is performed, model 4: selective type.

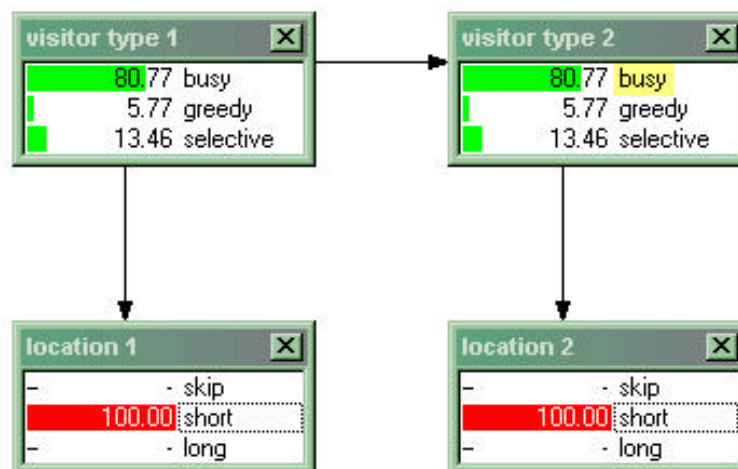


Figure 67. Posterior probabilities after evidence is introduced and inference is performed HMM-like model: busy type.



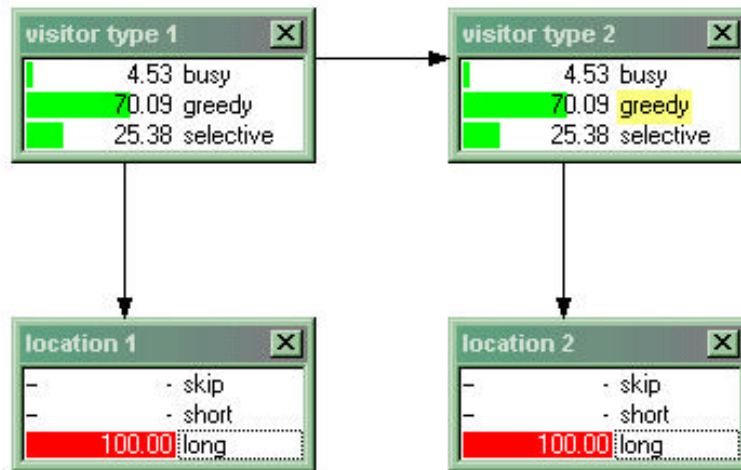


Figure 68. Posterior probabilities after evidence is introduced and inference is performed HMM-like model: greedy type.

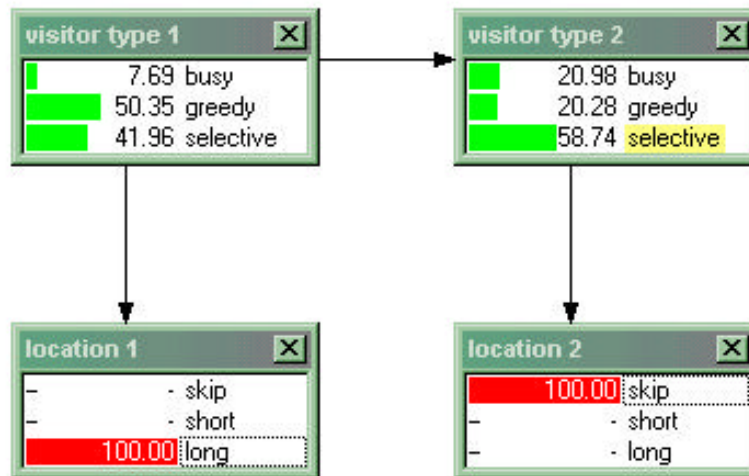


Figure 69. Posterior probabilities after evidence is introduced and inference is performed HMM-like model: selective type.



#### 4.2.5. Model Selection

The previous sections present five different Bayesian networks which all estimate the visitor type from the information provided by the location sensor. The models have been tested on all twelve time slices by introducing evidence by hand and comparing the obtained visitor highest probability with the visitor assigned label, in the tracking data. The figures above show only representative cases for two time slices, for the convenience of a printed document which allows the writer only to show a limited size/resolution picture. Rather than carrying out a full performance test for all five models (cross-validation), which would be quite a long procedure, it is desirable to be able to have other criteria that guide the selection of the best model amongst the ones proposed above. Certainly various criteria can be adopted. Speed is an important factor, as calculations of the visitor type with highest probability needs to be carried out in the real time in the museum wearable. Performance is another criteria, i.e. how accurately can the system guess the correct type. Given the limited number of objects that are usually present in a museum (maximum a few hundred), all the models proposed achieve fast calculation, therefore speed is not the main criteria in this case. As shown for the representative test cases with two time slices, all models have good performance and all lead to a plausible estimation of the user type. Other criteria are therefore needed to be able to select a model from the five proposed.

Model selection is still an open field of research for Bayesian networks [Hoeting, Madigan, Raftery, Volinsky, 1999] [Chickering and Heckerman, 1995] [MacKay, 1995]. MacKay provides a criteria for model comparison called the evidence framework [MacKay, 1992a, 1992b]. The general validity of such framework has recently been widely debated [Wolpert, 1993] and more recently questioned by Qi and Minka [Qi and Minka, 2001]. The agreement in the research community today is that yet the model's evidence provides at least a basic quantitative model raking criteria for Bayesian networks. Calculating the evidence for a Bayesian network can be in some cases a difficult task. Minka [Minka, 2001] proposes expectation propagation as a fast approximation technique for model selection.

For the purpose of this research, the evidence for the model  $H$ , given the data  $D$ :  $p(H|D)$  can be obtained as a byproduct of the sum-propagation probability update performed by Hugin, once evidence has been introduced for *all* the nodes of the network. The reason for this is that Hugin propagation is based on the junction-tree algorithm, and the operations Collect Evidence and Distribute Evidence. Once the messaging calls are finished, all probability tables are normalized so that they sum to one. The normalization constant for the root clique of the junction tree is the evidence of the model, given the data.

MacKay suggests using the evidence of the model as a criteria for comparison in the light of a Bayesian approach. According to MacKay, to evaluate the plausibility of two alternative hypothesis or models,  $H_1$  and  $H_2$ , we can use Bayes' theorem, to calculate the posterior probability for each of the model, given the observed data as:

$$p(H_i | D) = \frac{p(D | H_i) p(H_i)}{p(D)}$$

The quantity  $p(H_i)$  represents a prior probability for model  $H_i$ . If we have no particular reason to prefer one model over another, then we would assign equal priors to all of the models. Since the denominator  $p(D)$  does not depend on the model, we see that different models can be compared by evaluating  $p(D|H_i)$ .

This gives the following probability ratio between hypothesis (model)  $H_1$  and  $H_2$ :

$$\frac{p(H_1|D)}{p(H_2|D)} = \frac{p(H_1)}{p(H_2)} \frac{p(D|H_1)}{p(D|H_2)}$$

The first ratio  $\frac{p(H_1)}{p(H_2)}$  on the right hand side measures how much our initial beliefs favored  $H_1$  over  $H_2$ . The second ratio expresses how well the observed data was predicted by  $H_1$  compared to  $H_2$ .

This indicates that the Bayesian approach can be used to select a particular model for which the evidence is the largest. One might expect that the model with the greatest evidence is also the one which will have the best generalization performance. However MacKay shows us that that is only true, conditioning the model on the data, and that is why we consider  $p(H_i|D)$  as shown above. This is best illustrated by an example.

Let  $H_1, H_2, H_3$  be three different models with increasing flexibility, corresponding for instance to an increasing number of hidden units. Each model is given therefore by the network topology (number of units, conditional probabilities), and is governed by a number of active parameters. By varying the values of these parameters, each model can represent a range of cases. More complex models, with a greater number of hidden units, can represent a greater range of data sets. This is illustrated in figure 70.

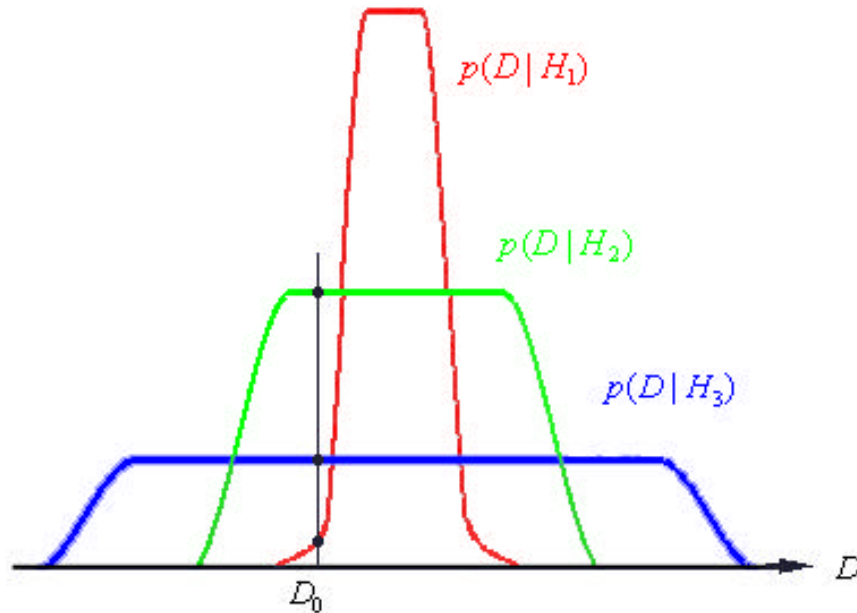


Figure 70. Comparison of model complexity, from MacKay, [MacKay, 1995].

When a particular data set  $D_0$  is observed, the model  $H_2$  has a greater evidence than either the simpler model  $H_1$  or the more complex model  $H_3$ .

According to MacKay, [MacKay, 1995] Bayesian model comparison embodies Occam's razor, the principle that states a preference for simple models. If several explanations are compatible with a set of observations, Occam's razor advises to go with the least complex explanation. The principle is often advocated both for aesthetical and practical reasons (empirical success). For MacKay, coherent inference embodies Occam's razor automatically and quantitatively.

Let see how these considerations apply to our problem. To rank the five models previously presented, I performed a simple test using the two time slice test described above. For each model, the data introduced is for test 1 is:  $D_1 = \{short, short, busy\}$ , for test 2 :  $D_2 = \{long, long, greedy\}$ , and for test 3:  $D_3 = \{long, skip, selective\}$ . To calculate the evidence for the model, the probability in all other nodes need to be set as well, and this is done based on domain knowledge. For example for model 1, if busy is set, then not\_greedy, and not\_selective can also be set. Similarly, not\_long, and not\_skip, are also set. Once all the known cases have been introduced in the network, in case of uncertainty, Hugin performs a max-propagation algorithm which selects the most likely combination of states given the evidence. The most likely states are then also set. Once all the states are set, the evidence for the network can be read as the normalization constant of the root clique of the junction tree. The results obtained are shown in table 17.

	test 1		test 2		test 3			
$p(H_i   D)$	t1	t2	t1	t2	t1	t2		
EVIDENCE	short	short	long	long	long	skip	product	ranking
model 1	0.002		0.004		0.0005		4 E-09	4
model 2	0.001		0.0034		0.001		3.4 E-09	5
model 3	0.005		0.0036		0.003		5.4 E-08	3
model 4	0.033		0.054		0.009		2 E-05	2
hmm	0.098		0.128		0.021		2.6 E-04	1

Table 17. Ranking of Bayesian networks according to the evidence framework

According to the evidence framework, the hmm-like Bayesian network is the highest ranked, followed by model 4a, followed by the others. What the above numbers also tell us is that the best model (hmm) is about fifty times better than the worst, and the best is also about two and a half times better than the second best. This is in accordance to Occam's razor: the models are ranked in order of complexity, from the most simple, the hmm, to the most complex. However model 4a, also ranks closely to the HMM, and even if slightly more complex, it also gives us information about the object, as it has, for each time slice, an added object node with a neutral/interesting/boring discrete state. Model 4a is therefore my final choice for estimating the visitor type as either busy, greedy, or selective, based on the sole stop duration information coming from the infrared location sensors.



## Chapter 5

# Sto(ry)chastics: editing stories for different visitor types and profiles

### 5.1. Content granularity and the knobs of a computational storytelling machine

The previous chapter has shown a Bayesian network based technique to identify the visitor's type in the museum. Rather than finding out what each individual wants, I have simplified the problem into classifying people's behavior at the museum in three basic types, according to the museum literature. This problem of profiling the user of a computer system is also known as *user modeling*. This chapter deals with content selection, conditioned on the user model, that is how to use the information obtained about the museum visitor to tailor a story which matches the visitor's intentions or desires. The intentions or desires are summarized and expressed by the visitor's type.

For a story, or message, to be conveyed, a human storyteller needs to have some knowledge about the narrated topic, and he/she needs to be able to articulate this knowledge into a form, which we usually call story. For example, a museum guide uses his/her knowledge about Van Gogh to explain to a group of visitors the particular style and meaning of a Van Gogh painting. Usually, a good storyteller will be able to articulate the story differently according to its audience. If the visitors are children from junior high school he/she will describe the artwork differently than if the visitors were tourists, or art experts. For the children he/she will simplify the language in which the story is told, and will make frequent stops and ask question to the children to make sure the audience is engaged. For the experts the storyteller will include many historical details, give precise information on the artistic practices and techniques, and will use a more sophisticated language. For the tourists he/she might enrich the description of the artwork with anecdotes to make it more entertaining. In certain occasions the museum guide will try and gather before or along the visit the visitors' preferences and will make longer stops, with longer and more articulated explanations next to the preferred artwork. If people don't seem interested the storyteller will instead cut short and move on to the next object on display.

From the above example, we can extract some basic, simplified, ingredients and recipe of storytelling which we can then use to parametrize a virtual storytelling machine hosted in

the museum wearable, and powered by the user modeling system described in the previous chapter.

The definition of story given below is quite simplistic, and nothing close to what one would find in a book about human narrative or storytelling. Yet computer programs are far from fully emulating certain high level human abilities, such as storytelling, and these simplifications are necessary to start building an interactive automatic storyteller.

The basic elements of story, for the purpose of this research, can be summarized as follows:

- 1. Knowledge about the topic: it can be summarized as the answers to: *who, what, when, where, why, how*.
- 2. Ability to articulate such knowledge into a story form.
- 3. Ability to adapt the story to the audience: i.e. ability to narrate a story with a different style or point of view, or slant (personalization: i.e. adapts to *what* the audience is interested in).
- 4. Ability to make the story long or short according to the level of interest shown by the public (adapts to *how much* the audience is interested in the narrated story, or wants to know about).
- 5. The story needs to make sense, and it possibly needs to be good.

A story narrated by a computer comes in the form of an audio-visual narration. In the specific case of this research, this narration is experienced through a mobile headphones/eye-glass display system that the user carries with them. Therefore the knowledge about a topic, for the museum wearable research, is represented by the collection of available audio-visual clips. The ability to articulate such knowledge into a story is similar to what in video or film is called sequencing. Sequencing is the task of assembling a variety of given shots into a visual narration. The museum wearable, as a computational storyteller, assembles together available audio-visual material, in such a way that the final cut is a compromise between the visitor's interests, what the curator says is a good story, and time which falls out of the user.

Here is how the above listed elements and parameters of story are mapped to parameters (knobs) of the automatic storyteller system for the museum wearable project:

- 1. Knowledge → described to the system in terms of content i.e. the collection of available video clips about the artwork on display.
- 2. Story articulation → real-time probabilistic reasoning about sequences of video clips.
- 3. Personalization → profiling: selection of the clips which best match the visitors' interests.
- 4. Overall length of presentation → the system should be able to tell stories which vary in size.

- 5. Sensemaking or good story → ordering of clips: the correct order of clips makes sure that cause-effect relationships are respected, and therefore that the story makes sense to its audience. Ordering is of course a very simplified way of thinking of good story and sensemaking. It is however a good first step in the framework of this research.

More specifically we need yet to define exactly what is *story form*, for this project, that is how to characterize the component elements that the system assembles together into a story. The building bricks of our system are characterized based on 1. their length and 2. their category. The length of the component elements in this context is not arbitrary, and it is determined by the actual length of a meaningful chunk of content. The problem of how to effectively segment the media into meaningful chunks is usually referred to, in the multimedia field, as the issue of *content granularity*. This is best explained with an analogy. If we had to assemble a written page from a book, we could choose whether our building bricks are words contained in the book, or at a higher level, full sentences of the book. This decision would have a high impact on the strategy we later need to implement to assemble the basic pieces together. Choosing building elements which are too small, such as words, has the advantage of producing a system which can be more flexible and creative in making a new assembly of words from the book. On the other hand the system needs to have a knowledge of grammar to be able to recombine the given words together into new sentences. Choosing formed sentences instead provides less flexibility, but it makes the subsequent editing task much easier. In this research the working assumption is that each story segment is a closed mini-narration with full meaning, and is to the whole story, what a sentence is to a paragraph of a written text.

The choice of categorizing the component elements into bins of related content material, is related to profiling. In the specific case of museums, most of the audio-visual material available for use by the museum wearable, tends to fall under a set of characterizing topics, which typically define art and science documentaries. This same approach to documentary as a composition of segments belonging to different themes, has been developed by Houbart in her work which edits a documentary based on the viewer's theme preferences, as an offline process [Houbart, 1994]. The difference between Houbart's work and what the museum wearable does is that the museum wearable performs editing in real time, using sensor input and Bayesian network modeling to figure out the user's preferences (type). After an overview of the audio-visual material available at MIT's Robots and Beyond exhibit, I identified the following content labels, or bins, I used to classify the component video clips.

*Story bins:*

- **Description** of the artwork: what it is, when it was created (answers: *when, where, what*)
- **Biography** of author: anecdotes, important people in artist's life (answers *who*)
- **History** of the artwork: previous relevant work of the artist
- **Context:** historical, what is happening in the world at the time of creation
- **Process:** particular techniques used or invented to create the artwork (answers *how*)
- **Principle:** philosophy or school of thought the author believes in when creating the artwork (answers *why*)

- **Form and Function:** relevant style, form and function which contribute to explain the artwork.
- **Relationships:** how is the artwork related to other artwork on display
- **Impact:** the critics' and the public's reaction to the artwork

This project required a great amount of editing to be done by hand (non automatically) in order to segment the two hours of video material available for the Robots and Beyond Exhibit at the MIT museum in the smallest possible complete segments. After this phase, all the component video clips were given a name, their length in seconds was recorded into the system, and they were also classified according to the list of bins described above. The classification was done probabilistically, that is each clips has been assigned a probability (a value between zero and one) of belonging to a story category. The sum of such probabilities for each clip needs to be one. The result of the clip classification procedure is shown in table 18.

I also conducted a study of how content is distributed geographically along the exhibit, both in two and three dimensions. The purpose of this study was to visualize the different stories for different visitors edited by the museum wearable as paths through the hyperspace of content in the exhibit. The 2D study shows colored pie charts in the vicinity of the twelve tracked objects at the museum. Each pie chart represents the content available for the corresponding object. The size of the pie chart is proportional to the amount of content available for that object. The size of the colored slices of the chart represent the contribution of each story bin to the content available for the object [figure 71]. I also leveraged some of my previous work [Sparacino, 1997] to provide a visualization of how the content bins contribute to create a storyscape specific to this exhibit. This time I used color coded vertical columns (a color for each content bin) whose height is proportional to the amount of content that each bin contributes to for the corresponding object. The results of this visualization study are shown in figures [72,73,74,75].

The museum wearable's storytelling system is an automatic real time sensor-driven and user-driven editing machine. Part of the success or good craft of the storytelling experience offered by the museum wearable resides in two phases which are not covered, or are beyond the scope of this research. The first phase concerns the production and construction of the story content, i.e. the original video material the museum provides to create the building blocks of story. The second phase is to do an accurate chunking of the available video into smaller units that are the smallest possible complete segments available for editing by the system. As of today, this editing work can best be done by a human. Recent research in automatic labeling of video database may provide grounding for future work which will allow the experience designer to extract automatically the labels for the content bins, and segment the video into smaller pieces belonging to these bins. However manual editing and labeling was required for this project. For all these reasons, the argument of this thesis is somewhat independent of the specific video clips available. Therefore in the rest of this document I will just use significant names to describe the component video segments used to test the system, as the specific content of each clip, once it has been labeled, does not contribute to prove the hypothesis of this work.



CATEGORIES / TITLES LENGTH IN SECS		Bitinside 021	Bitintro 090	Cogdrum 083	Cogfuture 043	Coghistory 051
Description	DSC	0.7	0.2	0.3	0	0
History	HST	0.1	0	0	0	0.5
Context	CTX	0	0	0	0	0
Biography	BIO	0.1	0.2	0	0	0.1
Process	PRC	0	0	0.6	0	0
Principle	PNC	0	0.4	0.1	1	0.2
Form & Function	FAF	0	0.2	0	0	0.2
Relationships	REL	0.1	0	0	0	0
Impact	IMP	0	0	0	0	0
Total P		1	1	1	1	1

CATEGORIES / TITLES LENGTH IN SECS		Cogintro 041	Dexdesign 114	Dexintention 034	Dexintro 072	Dexstiffness 096
Description	DSC	0.8	0.1	0.3	0.5	0.4
History	HST	0	0	0	0	0
Context	CTX	0	0	0	0.2	0
Biography	BIO	0	0.2	0	0	0
Process	PRC	0.2	0.3	0	0	0.2
Principle	PNC	0	0	0.2	0.1	0.4
Form & Function	FAF	0	0.4	0.5	0	0
Relationships	REL	0	0	0	0	0
Impact	IMP	0	0	0	0.2	0
Total P		1	1	1	1	1

CATEGORIES / TITLES LENGTH IN SECS		Itshort 026	Kiscynthiabo 067	Kisdevelop 066	Kisfacesensor 160	Kisintro 066
Description	DSC	0.8	0	0.1	0.3	0.3
History	HST	0	0	0.2	0	0
Context	CTX	0	0	0	0	0
Biography	BIO	0.2	0.9	0	0	0
Process	PRC	0	0	0.1	0.5	0.3
Principle	PNC	0	0	0.2	0	0.3
Form & Function	FAF	0	0	0.4	0.2	0.1
Relationships	REL	0	0.1	0	0	0
Impact	IMP	0	0	0	0	0
Total P		1	1	1	1	1

CATEGORIES / TITLES LENGTH IN SECS		Kissocial 183	Leg3Dbiped 055	Legblobby 021	Legflamingo 043	Legflamjerry 118
Description	DSC	0.1	0.3	1	0.5	0.4
History	HST	0	0	0	0	0
Context	CTX	0.1	0	0	0	0
Biography	BIO	0	0	0	0	0.2
Process	PRC	0.3	0.6	0	0.5	0.3
Principle	PNC	0.5	0	0	0	0.1
Form & Function	FAF	0	0	0	0	0
Relationships	REL	0	0.1	0	0	0
Impact	IMP	0	0	0	0	0
Total P		1	1	1	1	1

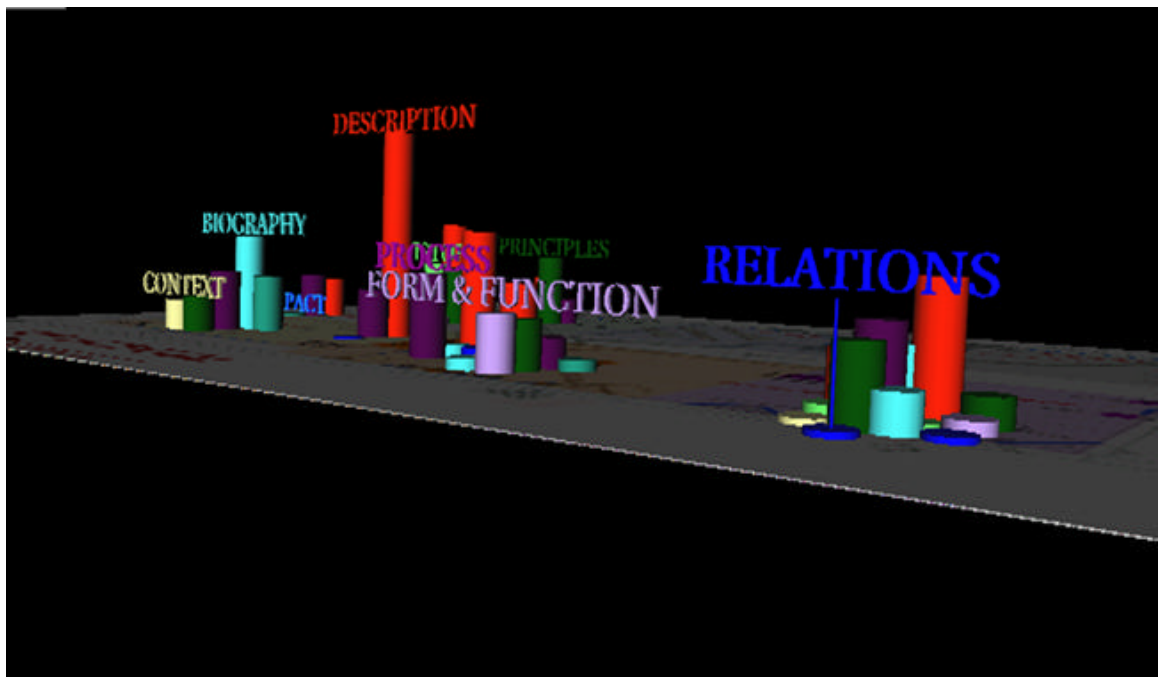
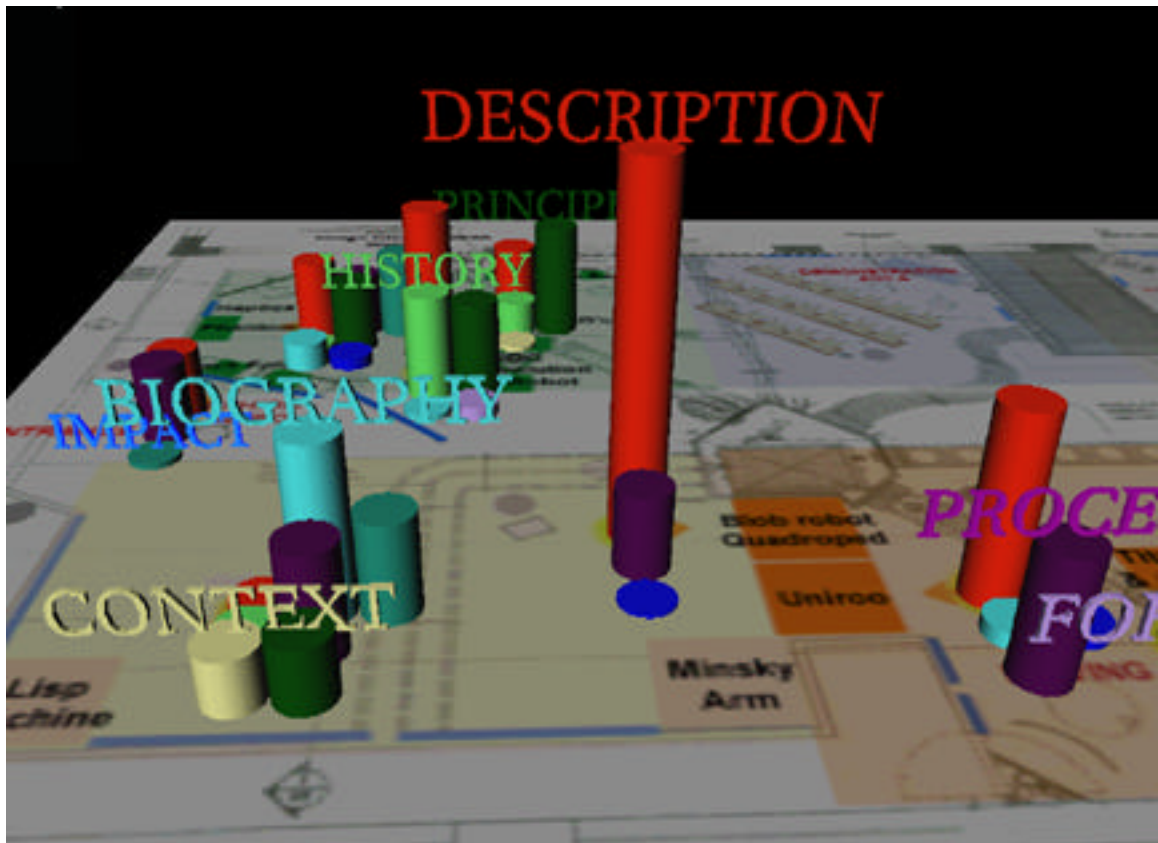
CATEGORIES / TITLES LENGTH IN SECS		Legmonopod 027	Legontherun 084	Legplbiped 054	Legquadropd 048	Leguniroo 017
Description	DSC	0.3	0.9	1	0.7	1
History	HST	0	0	0	0	0
Context	CTX	0	0	0	0	0
Biography	BIO	0	0	0	0	0
Process	PRC	0.7	0	0	0.25	0
Principle	PNC	0	0	0	0	0
Form & Function	FAF	0	0	0	0	0
Relationships	REL	0	0.1	0	0.05	0
Impact	IMP	0	0	0	0	0
Total P		1	1	1	1	1

CATEGORIES / TITLES LENGTH IN SECS		Phantappblin d 057	Phantappfree form 089	Phantapplacr oscp 074	Phantappsur g 120	Phanthowitw orks 041
Description	DSC	0.5	0	0.3	0.6	0.2
History	HST	0	0	0	0	0
Context	CTX	0	0	0	0	0
Biography	BIO	0	0	0	0	0
Process	PRC	0	0	0.7	0.3	0.8
Principle	PNC	0.4	0	0	0	0
Form & Function	FAF	0	0	0	0	0
Relationships	REL	0	0.2	0	0	0
Impact	IMP	0.1	0.8	0	0.1	0
Total P		1	1	1	1	1

CATEGORIES / TITLES LENGTH IN SECS		Phantintro 116	Kisemotions 168	Phanttombio 107
Description	DSC	0.3	0.5	0.2
History	HST	0.1	0	0
Context	CTX	0	0.1	0
Biography	BIO	0	0	0.4
Process	PRC	0.1	0.3	0
Principle	PNC	0.2	0.1	0.4
Form & Function	FAF	0	0	0
Relationships	REL	0	0	0
Impact	IMP	0.3	0	0
Total P		1	1	1

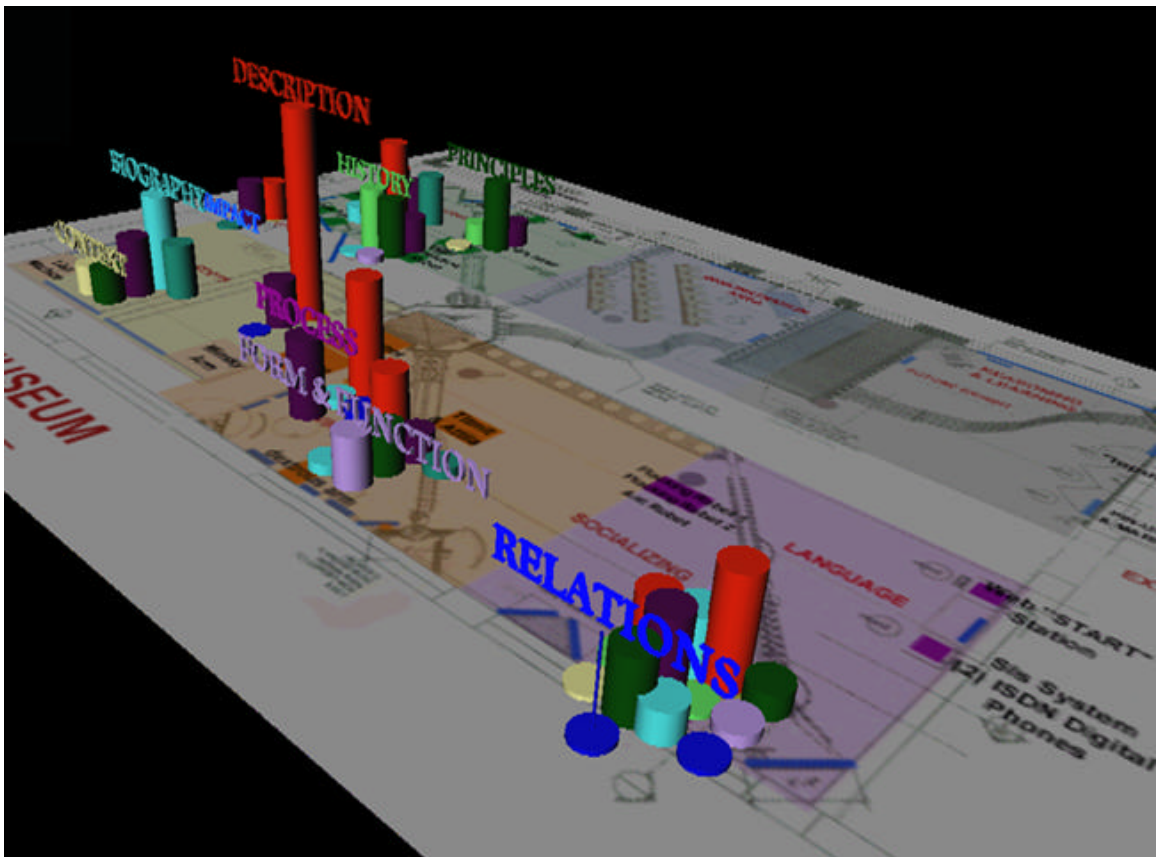
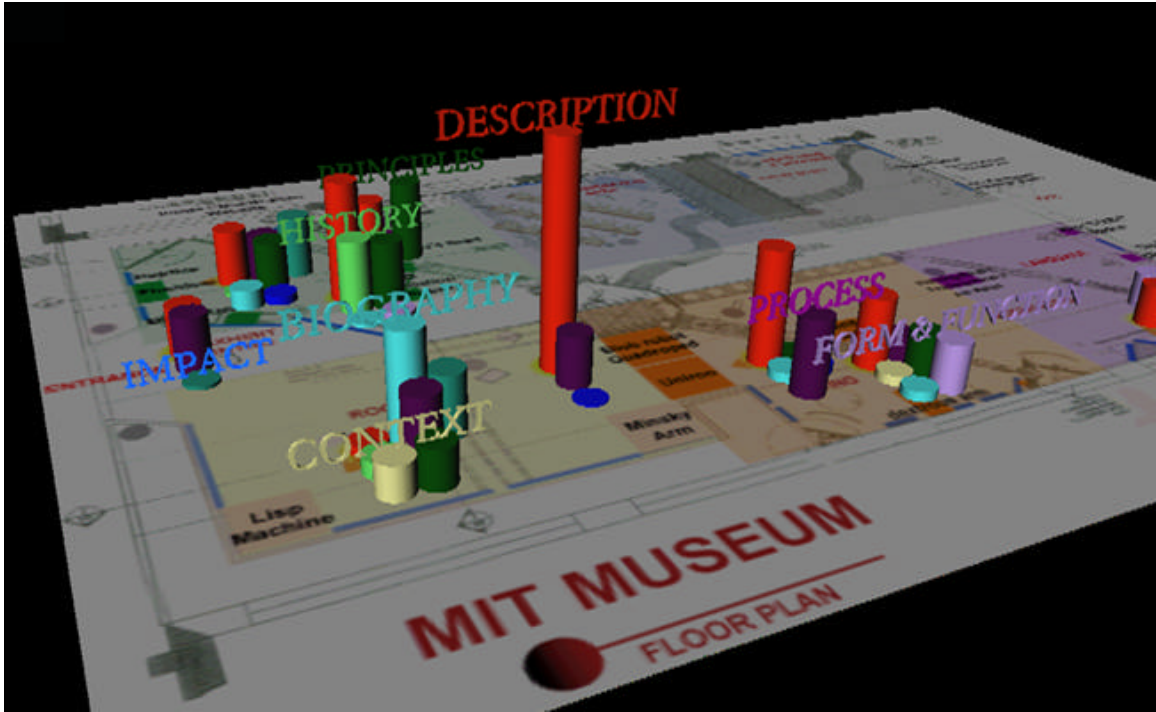
Table 18. Segments cut from the video documentation available for the MIT Museum's Robots and Beyond Exhibit. All segments have been assigned a set of probabilities which express their relevance with respect to nine relevant story themes or categories.





Figures 72, 73. Three dimensional representation of content distribution for MIT Museum's *Robots and Beyond* Exhibit.





Figures 74, 75. Three dimensional representation of content distribution for MIT Museum's *Robots and Beyond* Exhibit.

## 5.2. Content selection for different visitor types

Having described the knobs of a computational storytelling machine, this section illustrates how the museum wearable uses these knobs, with the knowledge of the visitor type obtained with a Bayesian network (Section 4.2), to assemble and sequence in real time a small audio-visual story relative to the object that the visitor is standing by. This should be seen as the first step towards a personalized user-driven and sensor-driven real time storytelling system. I further develop this discussion in section 5.3, which adds to the knowledge of the computational storytelling system not only the visitor type, but also the visitor interest profile, and therefore allows the system to articulate a more complex and personalized story for the visitor.

To perform content selection, “conditioned” on the knowledge of the visitor type, the system needs to be given a list of available clips, and the criteria for selection. There are two competing criteria in this case: one is given by the total length of the edited story for each object, and the other is given by the ordering of the selected clips. Therefore from the discussion in 5.1., *length* and *order* are the two story knobs which are used here for content selection. The order of story segments guarantees that the curator’s message is correctly passed on to the visitor, and that the story is a “good story”, in that it respects basic cause-effect relationships and makes sense to humans. Therefore in the Bayesian network which does content selection there will be a root node, called the “good story” node, which encodes, as prior probabilities, the curator’s preferences about how the story for each object should be told.

To make a decision about which clip(s) to play, the Bayesian network is extended to be an influence diagram: it will include decision nodes, and utility nodes which guide decisions. The decision node contains a list of all available content (movie clips) for each object. The utility nodes encode the two selection criteria: length and order. The utility node which describes length, contains the actual length in seconds for each clip. The length is transcribed in the network as a positive number, when conditioned on a preference for long clips (greedy and selective types). It is instead a negative length if conditioned on a preference for short content segments (busy type). This is because a utility node will always try to maximize the utility, and therefore length is penalizing in the case of a preference for short content segments. The utility node which describes order, contains the profiling of each clips into the story bins described in section 5.1. and listed in table 18 times a multiplication constant used to establish a balance of power between “length” and “order”. Basically order here means a ranking of clips based on how closely they match the curator’s preferences expressed in the “good story” node. The selection of the first and subsequent content segments is therefore a function the clip’s length, and of how closely it matches the curator’s preferences. By means of probability update, the Bayesian network comes up with a “compromise” between length and order and provides a final ranking of the available content segments in the order in which they should be played. The network from the previous chapter is extended to do content selection in addition to visitor type identification [figure 76], and the priors, utility, and decision nodes are given in tables 25,26,27,28.

This network, shows for simplicity a situation in which only one optimal content segment is selected. However through the ranking of content segments provided by the decision nodes for each object, the network can actually select more than one segment, as a compromise between order of preference and story length. In the network shown in figure 76 the number of clips which are played is limited by the overall story length available for each visitor type. This means that the type has an influence not only on the decision of which segments are played, based on their length, but also in the number of segments that are played, based on their *total* length. Before content assembly, the curator and the system modeller need to establish a maximum length of story in seconds for each object and each type such as the one in table 19:

		max length obj 1	max length obj 2
<b>BUSY</b>	<b>a little of everything</b>	<b>90</b>	<b>90</b>
<b>GREEDY</b>	<b>a lot of everything</b>	<b>480</b>	<b>220</b>
<b>SELECTIVE</b>	<b>a lot of the same</b>	<b>480</b>	<b>220</b>

Table 19. Maximum duration of story (sum of segments) for the three visitor types

From this table the reader should notice that while the greedy and selective type are given the same maximum allowed story length for an object, the system uses a different criteria to concatenate segments for the two types. According to the simplified and somewhat stereotypical type definition given in section 3.3., that serves as a working hypothesis for this research, while a greedy type wants to see a lot of everything, the selective type will want to see a lot only within the themes that he/she is most interested in. Not only will these two types get a different story, but neither will see all the material that is available to the system for each object. Usually the museum has usually a lot more audiovisual information that anybody can browse in the time dedicated to a museum visit. As described in section 3.2, the museum wearable assembles a short (2 minutes) to a maximum length (10 minutes) stories for each object from the available audiovisual database of several hours. If the wearable were to show all of its available content for one object to the greedy type, his/her visit could last as long as 6 to 10 hours, which is inconceivable (in one day) even for the most motivated visitor. Extension of the system to be personalized for repetitive use across several days is beyond the scope of the research described in this document, but is considered a desirable extension of the system for future work. The calculation of which segments are played for each visitor stop is:

do {

2. select highest ranked clip (in order of individual segment's length, and curator's preferences which express ordering of a good story)
3. select next ranked clip. For the selective type, select next clip only if similar to the previously selected. If not, stop.
4. check if the overall length of story is less than the maximum allowed for each type/object (as in point 1).
5. If so, keep this segment, and find the next one: go to #3.
6. If not, abandon the current segment selection, and try another one: go to #3.

} while there are still content segments available for the current object. Then stop.

Note that the additional nodes in figure 77 are shown only to explain that there is an additional computation to calculate the number of segments to be assembled. This computation is actually performed in software, using the ranking provided by the Bayesian network, but using C++ instructions instead, inside the program that handles both the network and the content playout (section 6.4). The reason is that it would be cumbersome to calculate with the network a simple algorithm as the one described above, which can be easily performed in any computer language, but not so easily with a Bayesian network (it's a deterministic calculation, and not a probabilistic one).

Tables 29 and 30 show the results of content selection when two different definitions (preferred segment ordering) are given to the network. This is done to show how the curator's viewpoint or message can be easily taken into account in this framework by simply setting (or changing) the prior probabilities in the "good story" node. Choosing a good order of story bins to produce an arrangement of segments which cognitively makes sense to the visitor, is something important which is reflected in the prior probabilities of the "good story" node. If the segments are well cut, so that they encapsulate a minimum length, yet complete, mini-story, and if the priors are well chosen, the resulting story will make sense to the human viewer. Particular care needs therefore to be taken in these two preparatory steps for the system, which are done by humans by hand. No matter how good the network, and the content selection mechanism, if the original segments are badly shot, and cut in a sloppy way, and the priors are not thoughtfully chosen, the story produced by the system could be disjointed and non compelling.

Story bins or CATEGORIES		curator Frank	Curator Liz
Description	DSC	0.3	0.1
History	HST	0.1	0.19
Context	CTX	0.03	0.03
Biography	BIO	0.16	0.21
Process	PRC	0.14	0.02
Principle	PNC	0.06	0.2
Form & Function	FAF	0.09	0.12
Relationships	REL	0.08	0.05
Impact	IMP	0.04	0.08
Total P		1	1

Table 20. Two different views of what is a "good story" according to curator Frank and curator Liz.

Table 20 shows to possible definition of "good story", in the framework of this research, by two different curators, called for easy reference, Frank and Liz. What the numbers above say is that Frank believes that a good museum story should start with an extensive object description, followed by biographical information about its creator. Next, explanation about the process of creation should be given, accompanied by a history of previous versions or sketches of the same object, and elements of form and function. Of less importance are the relationship of the object to other objects on display, the guiding philosophical principles which have led to its creation, its impact of the public and the art critics, and what was happening in the world at the time of creation. Liz thinks differently than Frank. She believes that a good museum story should be based on the creator's profile and biographical information, and that these elements should have the priority.



Explaining to the public what previous artwork has led to the creation of the object they are looking at is also important. Information about Form and Function, accompanied by a more detailed description of the object should follow. All the other themes, or story bins, are of secondary importance. The different results of these two rather opposite views of story are shown in tables 29 and 30. These possible choices of Liz and Frank are given only to provide examples how the system works. The same curator could actually choose a different combination of weights for the good story node, for a different exhibit, as the message carried by an exhibits changes with its content.

To provide examples of content selection I present three test cases, for the busy, greedy, and selective types [figures 78-83], with the good story prior probabilities set by curator Frank. I then show the different selection that the system makes with the priors, given by curator Liz [figures 84-89]. For limited space on paper, these test cases are presented only for the first two objects on display.

**Test case 1.** The visitor spends a short time both with the first and second object → the network gives the highest probability to the *busy* type.

**Test case 2.** The visitor spends a long time both with the first and second object → the network gives the highest probability to the *greedy* type.

**Test case 3.** The visitor spends a long time with the first object and a short time with the second object → the network gives the highest probability to the *selective* type.

Model parameters:

Transition probabilities p(O2 O1)			
	neutral	interesting	boring
neutral	0.85	0.075	0.075
interesting	0.075	0.85	0.075
boring	0.075	0.075	0.85

Preferred length			
	busy	greedy	selective
short	1	0	0.2
long	0	1	0.8

Cog clips	length
Play cogdrum	83
Play cogfuture	43
play coghistory	51
play cogintro	41

Kismet clips	length
Play kiscynthiablo	67
Play kisdevelop	66
Play kisfacesensors	160
Play kisintro	66
Play kissocial	183
Play kisemotions	168

length	Long					
kismet clips	play kiscynthiablo	play kisdevelop	Play kisfacesensors	play kisintro	play kissocial	play kisemotions
	67	66	160	66	183	168
length	Short					
kismet clips	play kiscynthiablo	play kisdevelop	Play kisfacesensors	play kisintro	play kissocial	play kisemotions
	-67	-66	-160	-66	-183	-168

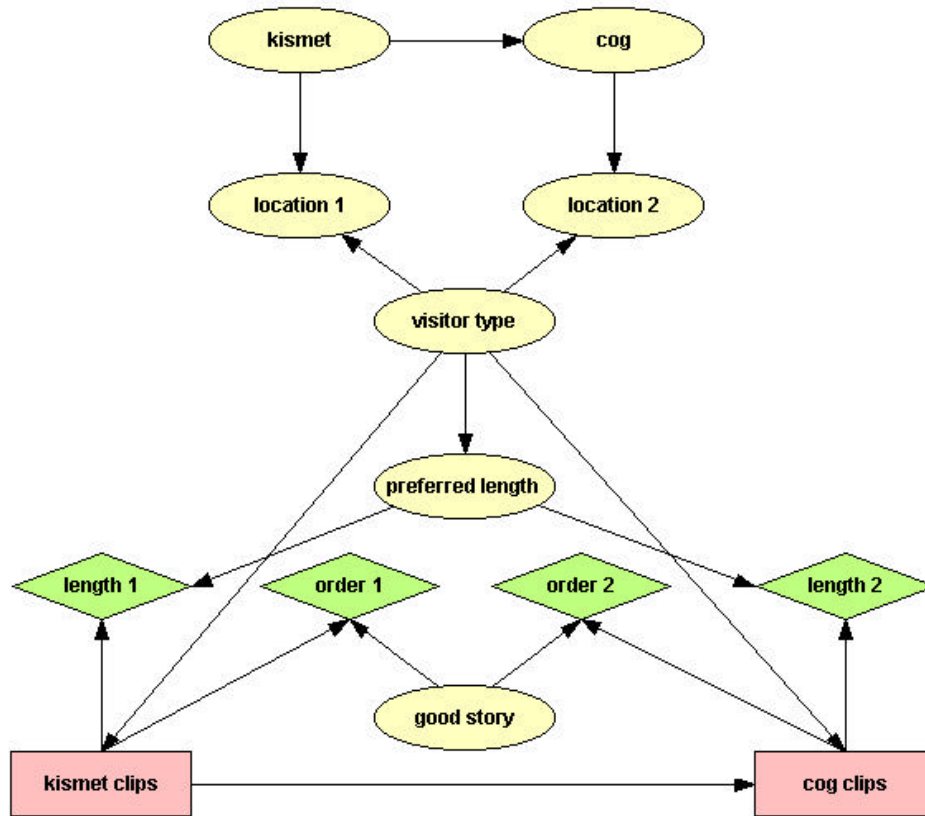
Tables 21,22,23,24,25. Model parameters for the Bayesian network for content selection

preferred length	Long			
cog clips	play cogdrum	play cogfuture	play coghstory	play cogintro
	83	43	51	41
preferred length	Short			
cog clips	play cogdrum	play cogfuture	play coghstory	play cogintro
	-83	-43	-51	-41

kismet clips	Play kiscynthiabi								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
				270				30	
kismet clips	Play kisdevelop								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	30	60			30	60	120		
kismet clips	Play kisfacesensors								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	90				150		60		
kismet clips	Play kisintro								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	90				90	90	30		
kismet clips	Play kissocial								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	30		30		90	150			
kismet clips	Play kisemotions								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	150		30		90	30			

cog clips	Play cogdrum								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	90				180	30			
cog clips	Play cogfuture								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
						300			
cog clips	Play coghstory								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
		150		30		60	60		
cog clips	Play cogintro								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	240				60				

Tables 26,27,28. Model parameters for the Bayesian network for content selection  
Tables 27, 28 include the same parameters of Table 18 (pg 88,89) multiplied by a weighting factor of 300



Figures 76, 77. Extension of Bayesian network for visitor type identification to content selection.

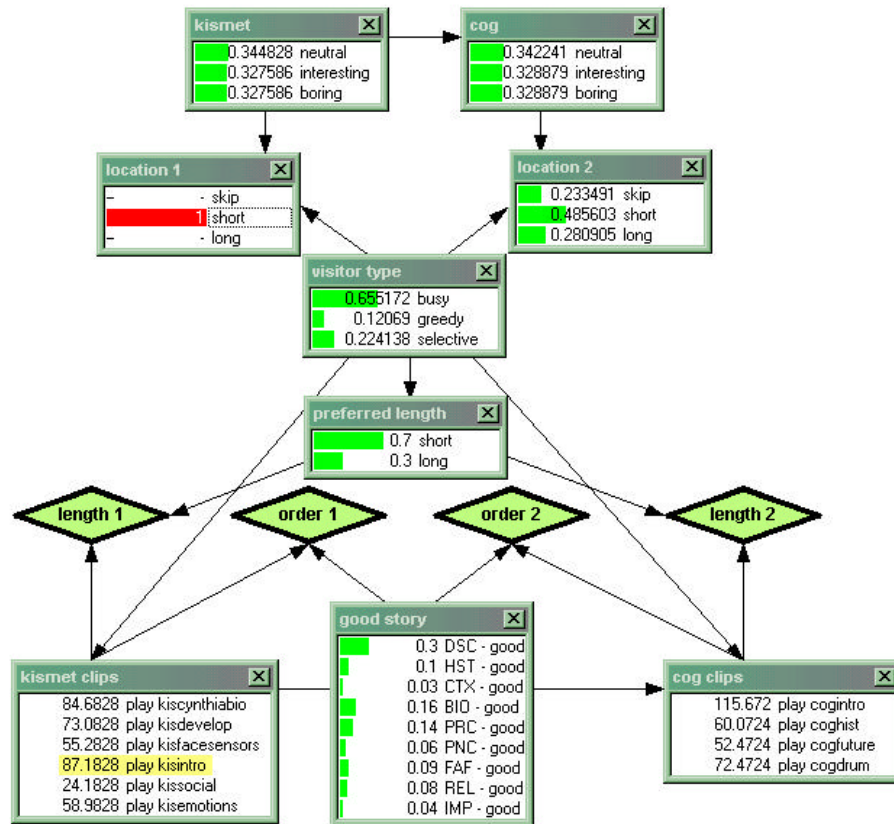


Fig.78

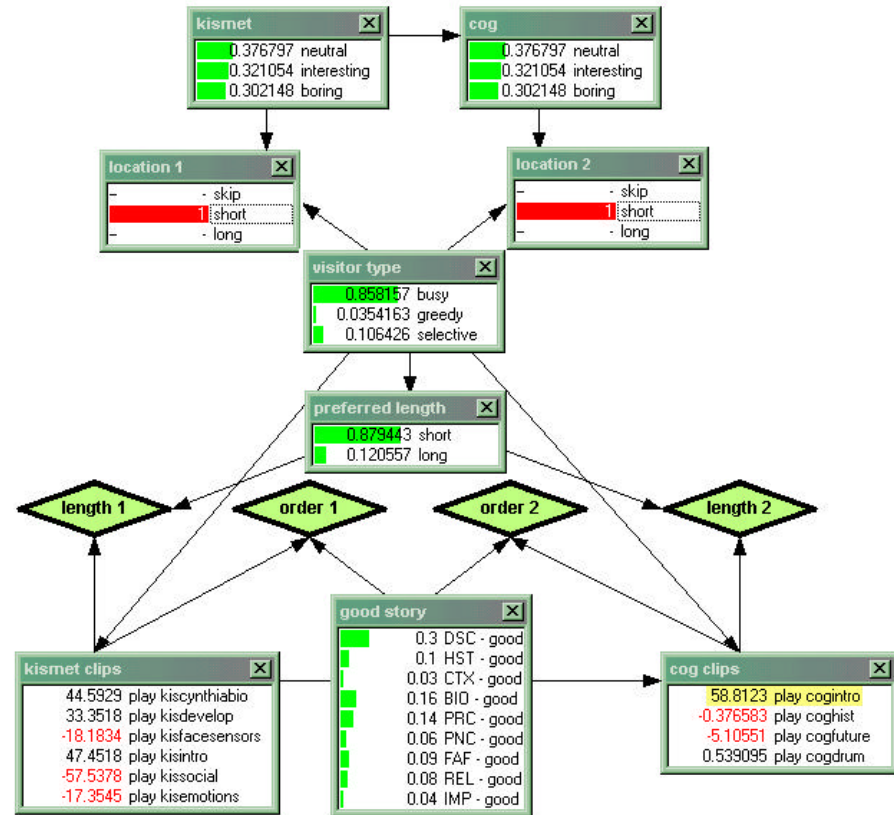


Fig.79

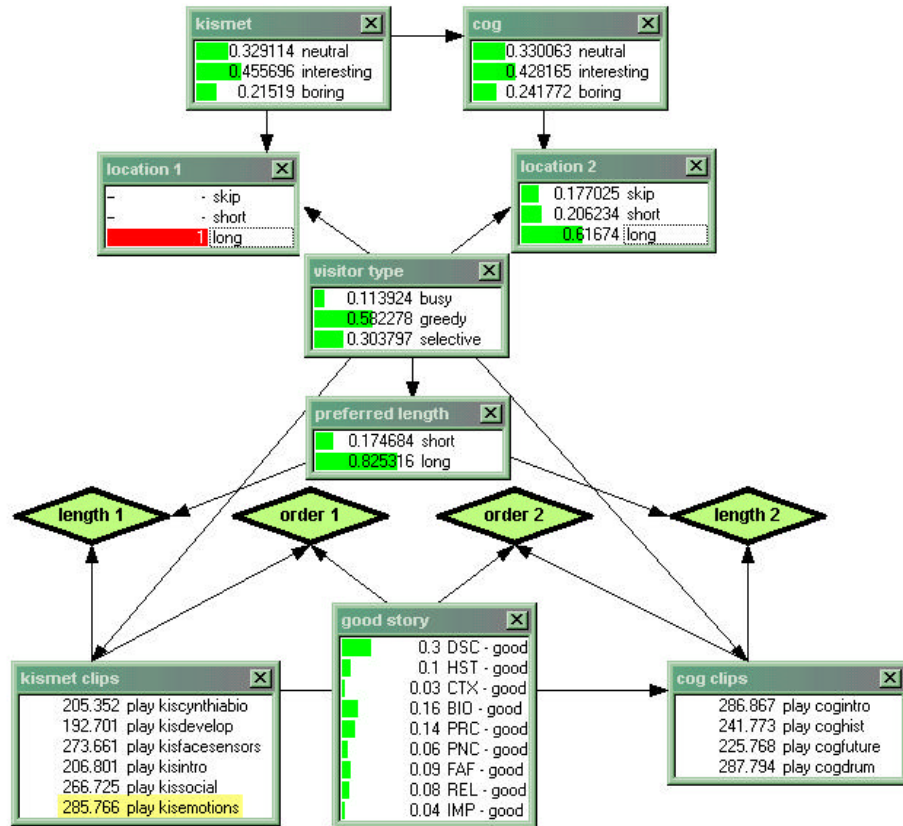


Fig.80

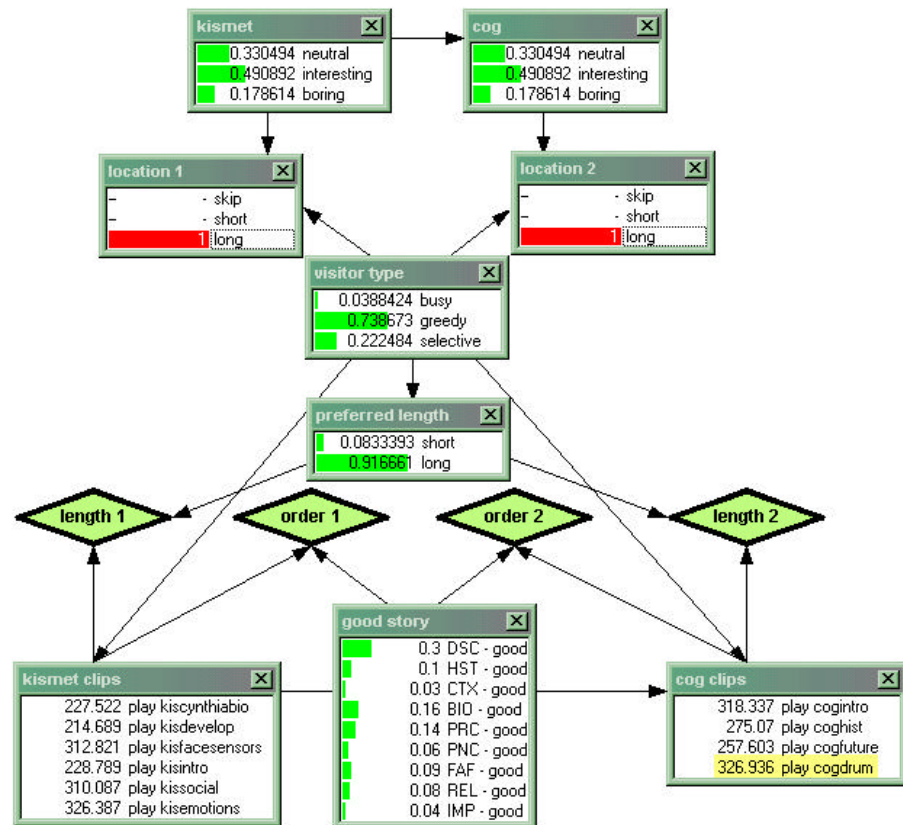


Fig.81

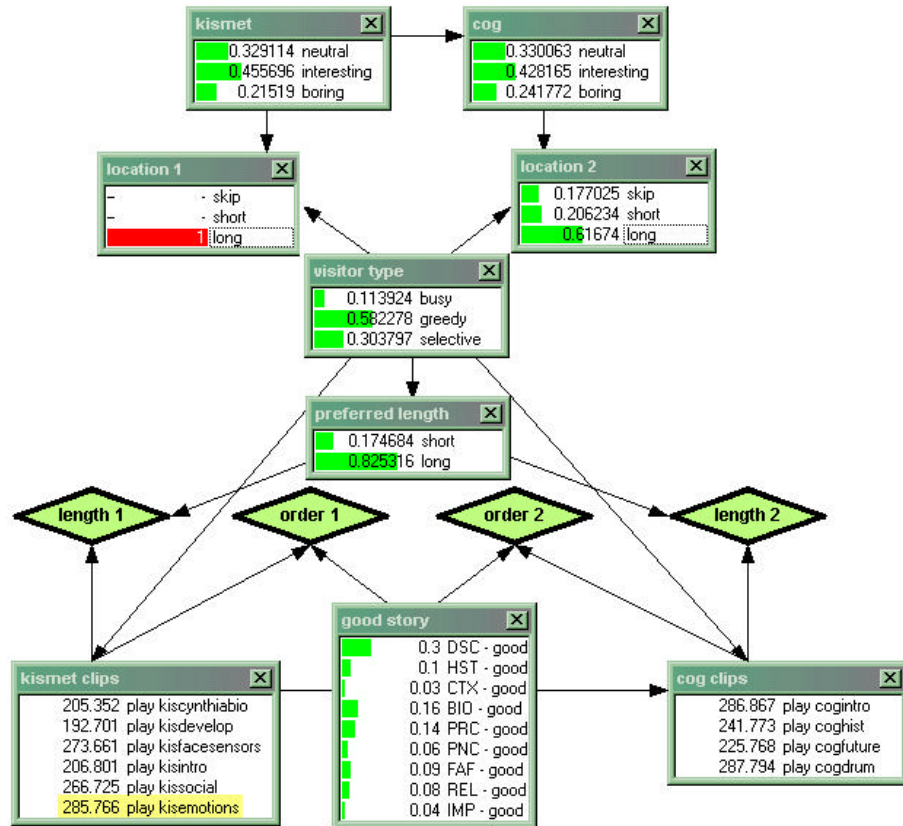


Fig.82

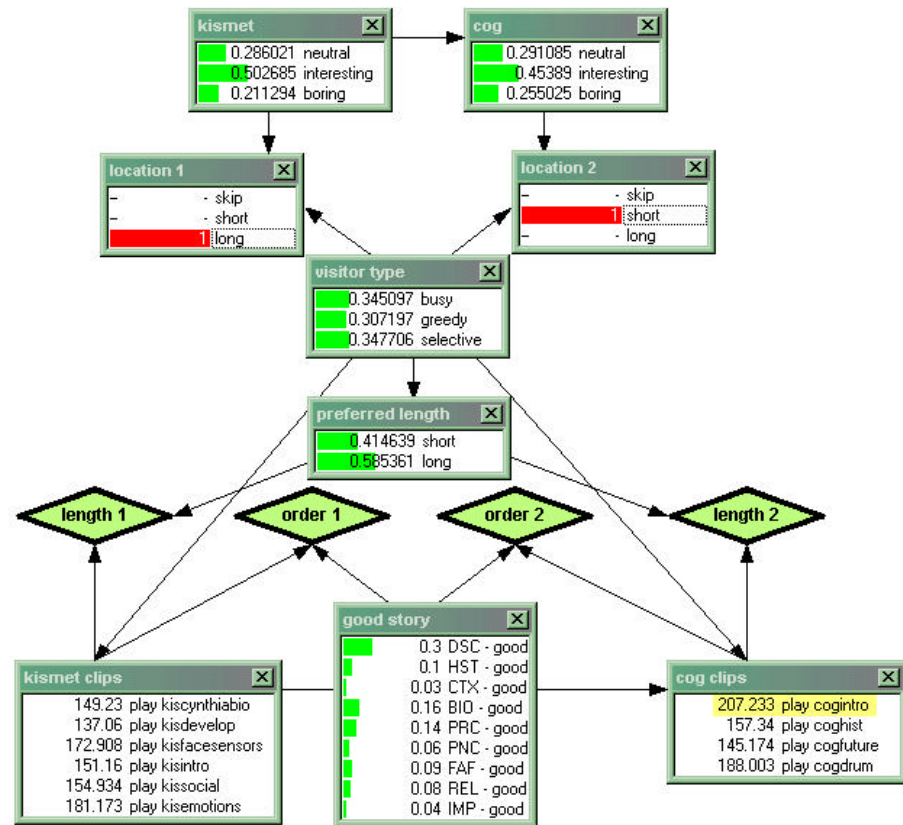


Fig.83

The results for the three test cases in figures 78-83 are summarized below. The first column shows the content selection in case the system was to show only one clip per object. The remaining columns, and the rows highlighted in red, show how the system edits a small story for each object, based on the description above. The reader should notice how the system selects for the busy type the smallest description clip, called “intro” for both objects, the Kismet and Cog robots on display at MIT’s Robots and Beyond Exhibit. If more than one clip is selected the museum wearable also adds the *cogfuture* segment to the story played for the cog object, as it is the next available short clips which fits within the 90 seconds maximal story duration available for the busy type. For the greedy type the system gives highest ranking to the *kisemotions* clip and the *cogrum* clip, which both have high weights on the description and process themes, preferred by the curator for this hypothetical museum wearable augmented exhibit. The selective type sees the cogintro clip at his/her second short stop, which is coherent with the specifications of the system. All these cases demonstrate how the museum wearable selects content appropriately and in accordance to the visitor type profiling and corresponding expectations.

#1	Table 29. Content segment assembly based on story and segment length									
test 1	Short									
selected	Clip 1	length 1	clip 2	length 2	total time					
busy	Kisintro	66	kiscynthiabio	67	133					
	Kisintro	66	kisdevelop	66	132					
	Kisintro	66	kisemotions	168	234					
	Kisintro	66	kisfacesensors	160	226					
	Kisintro	66	kissocial	183	249					
	kisintro	66			66					
	Short									
	Clip 1	length 1	clip 2	length 2	clip 3	length 3	clip 4	length 4	total time	
	Cogintro	41	cogdrum	83	coghist	51	cogfuture	43	218	
	Cogintro	41	coghist	51					92	
	Cogintro	41	cogfuture	43					84	
	cogintro	41	cogfuture	43					84	
test 2	Long									
selected	Clip 1	length 1	clip 2	length 2	clip 3	length 3	total time			
greedy	Kisemotions	168	kisfacesensors	160	kissocial	183	511			
	Kisemotions	168	kisfacesensors	160	kisintro	66	394			
	kisemotions	168	kisfacesensors	160	kisintro	66	394			
	Long									
	Clip 1	length 1	clip 2	length 2	clip 3	length 3	clip 4	length 4	total time	
	cogdrum	83	cogintro	41	coghist	51	cogfuture	43	218	
test 3	Long									
selected	Clip 1	length 1	clip 2	length 2	clip 3	length 3	total time			
selective	Kisemotions	168	kisfacesensors	160	kissocial	183	511			
	kisemotions	168	kissocial	183			351	similar segments		
	Short									
	Clip 1	length 1	clip 2	Length 2	clip 3	length 3	clip 4	length 4	total time	
	Cogintro	41	cogdrum	83	coghist	51	cogfuture	43	218	
	cogintro	41	cogdrum	83	coghist	51			175	



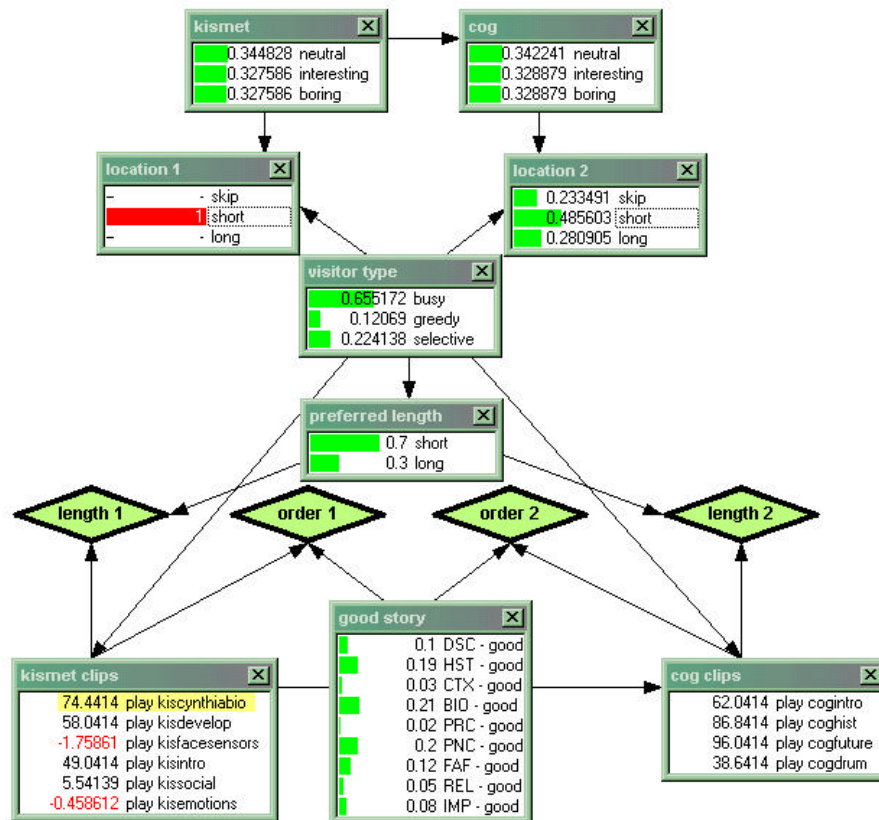


Fig.84

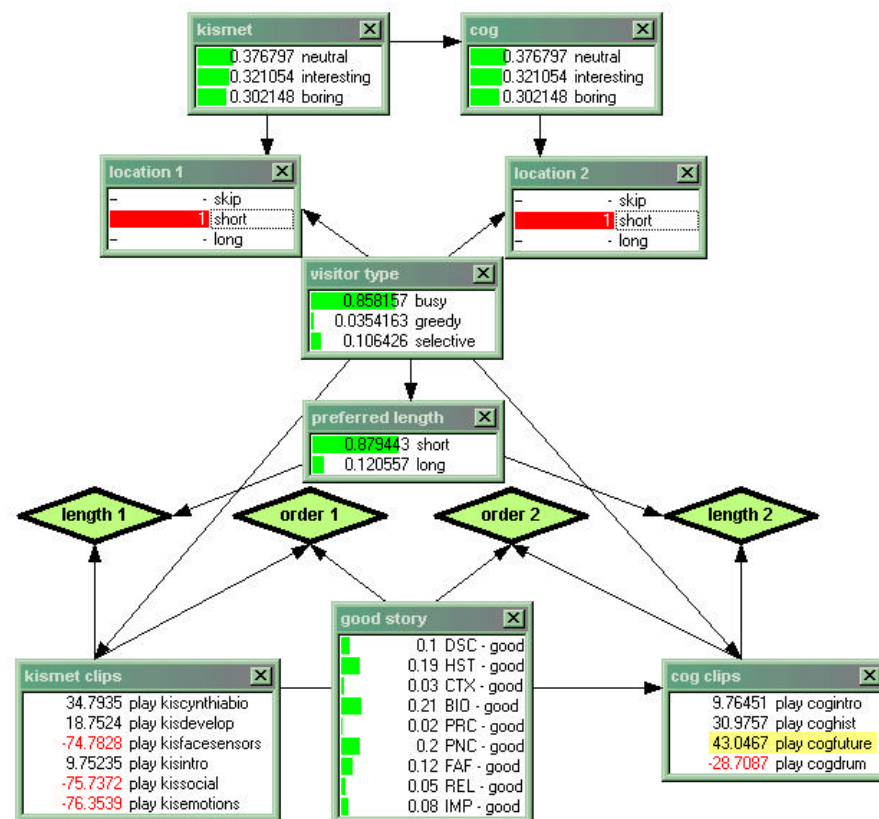


Fig.85



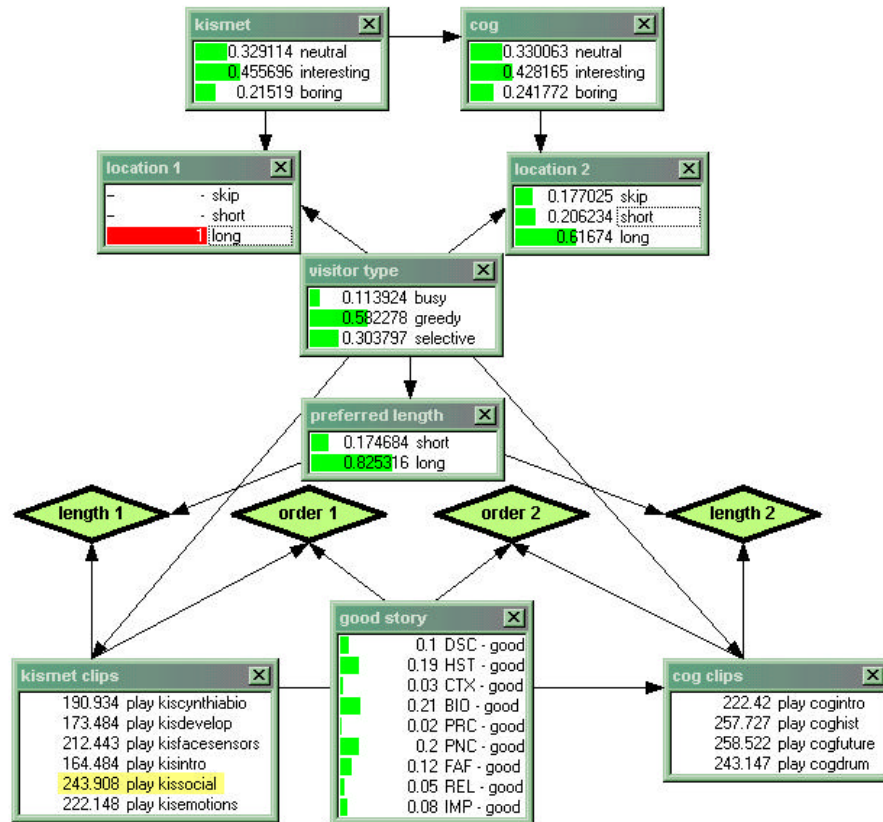


Fig.86

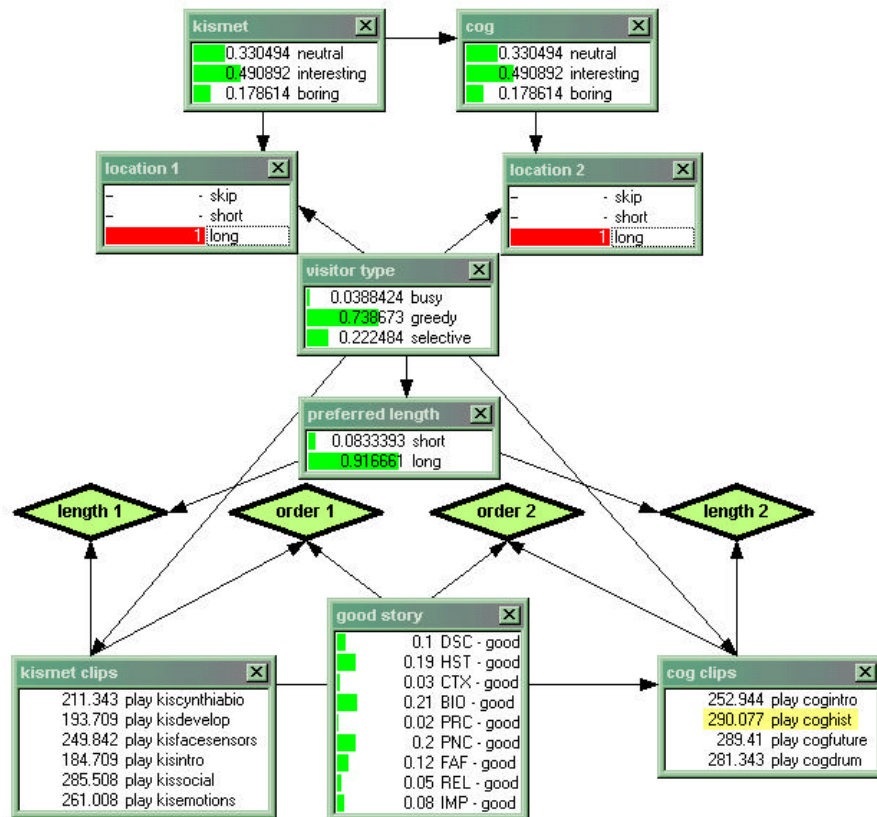


Fig.87

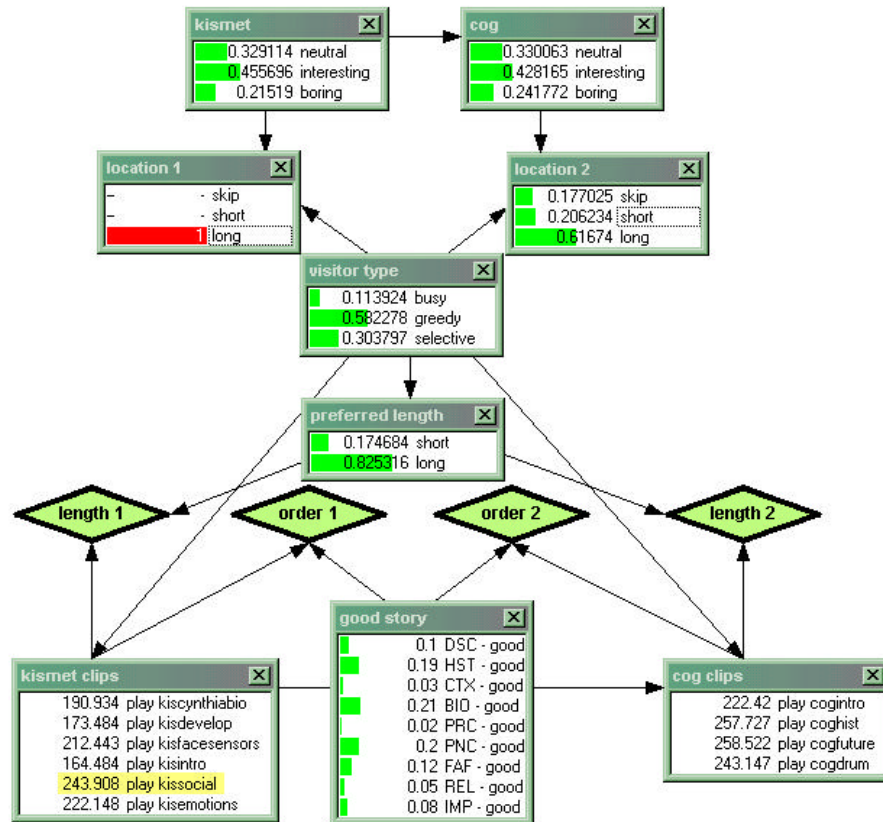


Fig.88

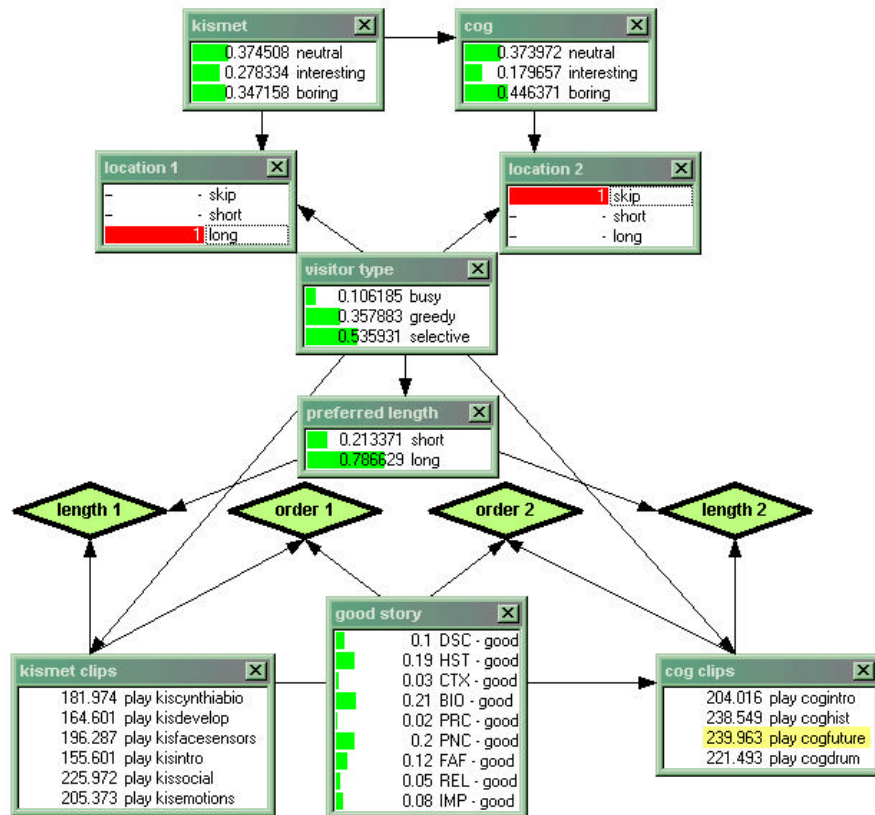


Fig.89

#2	Table 30. Content segment assembly based on story and segment length, different "good story" node than Table 29.									
test 1	short									
selected	clip 1	length 1	clip 2	length 2	total time					
busy	kiscynthiablo	67	kisdevelop	66	133					
	kiscynthiablo	67			67					
	short									
	clip 1	length 1	clip 2	length 2	clip 3	length 3	clip 4	length 4	total time	
	cogfuture	43	coghist	51						94
	cogfuture	43	cogdrum	83						126
	cogfuture	43	cogintro	41						84
test 2	long									
selected	clip 1	length 1	clip 2	length 2	clip 3	length 3	total time			
greedy	kissocial	183	kisemotions	168	kisfacesensors	160	511			
	kissocial	183	kisemotions	168	kiscynthiablo	67	418			
	long									
	clip 1	length 1	clip 2	length 2	clip 3	length 3	clip 4	length 4	total time	
	coghist	51	cogfuture	43	cogdrum	83	cogintro	41		218
	coghist	51	cogfuture	43	cogdrum	83				177
test 3	long									
selected	clip 1	length 1	clip 2	length 2	clip 3	length 3	total time			
selective	kissocial	183	kisemotions	168	kisfacesensors	160	511			
	kissocial	183	kisemotions	168			351	similar segments		
	short									
	clip 1	length 1	clip 2	length 2	clip 3	length 3	clip 4	length 4	total time	
	cogfuture	43	coghist	51	cogdrum	83	cogintro	41		218
	cogfuture	43	coghist	51	cogdrum	83				177

Table 30 shows how the system selects different clips if a different preferred ordering is specified in the good story node. At this point, a third curator, let say Nancy, could argue that, in her view, a story should always start with the shortest possible introductory and descriptive clip, followed by a longer description clip, followed by a segment which describes the creative process and so on. Her preferences can be easily accomodated by having the system always select the shortest description clip in first place, and then using the segment ranking provided by the Bayesian network for the following segments. The Bayesian network leaves therefore plenty of choice to the exhibit designer, curator, and storyteller, on the preferred story editing ordering and criteria. What it provides them is easy access to the knobs of the virtual storytelling machine described in section 5.1. without the need to calculate in advance all possible combinations given by all the knob values. A more in depth comparison between Bayesian networks and traditional multimedia systems, based one one to one mappings between sensors (inputs) and content (outputs) is described in section 7.2.

An alternative to clip selection, other than the one shown in tables above, consists in replicating the decision node for each segment selection, and having a transition node

which expresses preferences in segment concatenation. Table 31 shows the preferred theme transitions for the stories told by the museum wearable. These transitions are set as prior probabilities and have only heuristic value. The choice of which theme follows which, is up to the curator, the content designer, and the system modeller, and therefore the table below [table 31] shows a subjective preference chosen for this project. Another team of designers may have as well chosen other theme transitions. Whatever the choice, what the Bayesian network does for us is to give us a means to express this choice into a node which then conditions the choice of the subsequent segment.

<b>Description → Form and Function</b> <b>Description → Impact</b> <b>Description → Relations</b> <b>Descriptions → Biography</b> <b>Descriptions → Principles</b>	<b>History → Description</b> <b>History → Form and Function</b> <b>History → Principles</b> <b>History → Process</b> <b>History → Relations</b>
<b>Context → Biography</b> <b>Context → Description</b> <b>Context → Form and Function</b> <b>Context → Impact</b> <b>Context → Principles</b> <b>Context → Process</b> <b>Context → Relations</b>	<b>Principles → Description</b> <b>Principles → Context</b> <b>Principles → Form and Function</b> <b>Principles → Implications</b> <b>Principles → Process</b>
<b>Process → Descriptions</b> <b>Process → Form and Function</b> <b>Process → Impact</b> <b>Process → Principle</b>	<b>Relations → Impact</b> <b>Relations → History</b> <b>Relations → Context</b>
<b>Biography → Principles</b>	<b>Form and Function → Implications</b>
<b>Impact → Biography</b> <b>Impact → Principles</b>	

Table 31. Table with a possible example of segment sequencing constraints

An example of this technique is shown in figure 90. This new network, illustrated for simplicity only for the first object, uses the “play next” node to encode the transition probabilities below. A long stop duration at object one, with the “good node” as in case 1, originally gave a story made by the segments: [*kisemotions*, *kisfacesensors*, *kisintro*] [table 29]. With his technique instead, the final story has a slightly different composition [figures 91-93]: [*kisemotions*, *kissocial*, *kisintro*]. Note that the last clip is the second best choice, because it is of a shorter length than the best choice for that segment, and therefore fits within the maximum allotted playout time for the greedy type. Below [table 32] is the conditional probability table for the segment transition node:

Story themes		DSC	HST	CTX	BIO	PRC	PNC	FAF	REL	IMP
<b>Description</b>	<b>DSC</b>	0	0.2	0.143	0	0.25	0.2	0	0	0
<b>History</b>	<b>HST</b>	0	0	0	0	0	0	0	0.333	0
<b>Context</b>	<b>CTX</b>	0	0	0	0	0	0.2	0	0.333	0
<b>Biography</b>	<b>BIO</b>	0.2	0	0.143	0	0	0	0	0	0.5
<b>Process</b>	<b>PRC</b>	0	0.2	0.143	0	0	0.2	0	0	0
<b>Principle</b>	<b>PNC</b>	0.2	0.2	0.143	1	0.25	0	0	0	0.5
<b>Form &amp; Function</b>	<b>FAF</b>	0.2	0.2	0.143	0	0.25	0.2	0	0	0
<b>Relationships</b>	<b>REL</b>	0.2	0.2	0.143	0	0	0	0	0	0
<b>Impact</b>	<b>IMP</b>	0.2	0	0.143	0	0.25	0.2	1	0.333	0

Table 32. Conditional probability table for the segment transition node

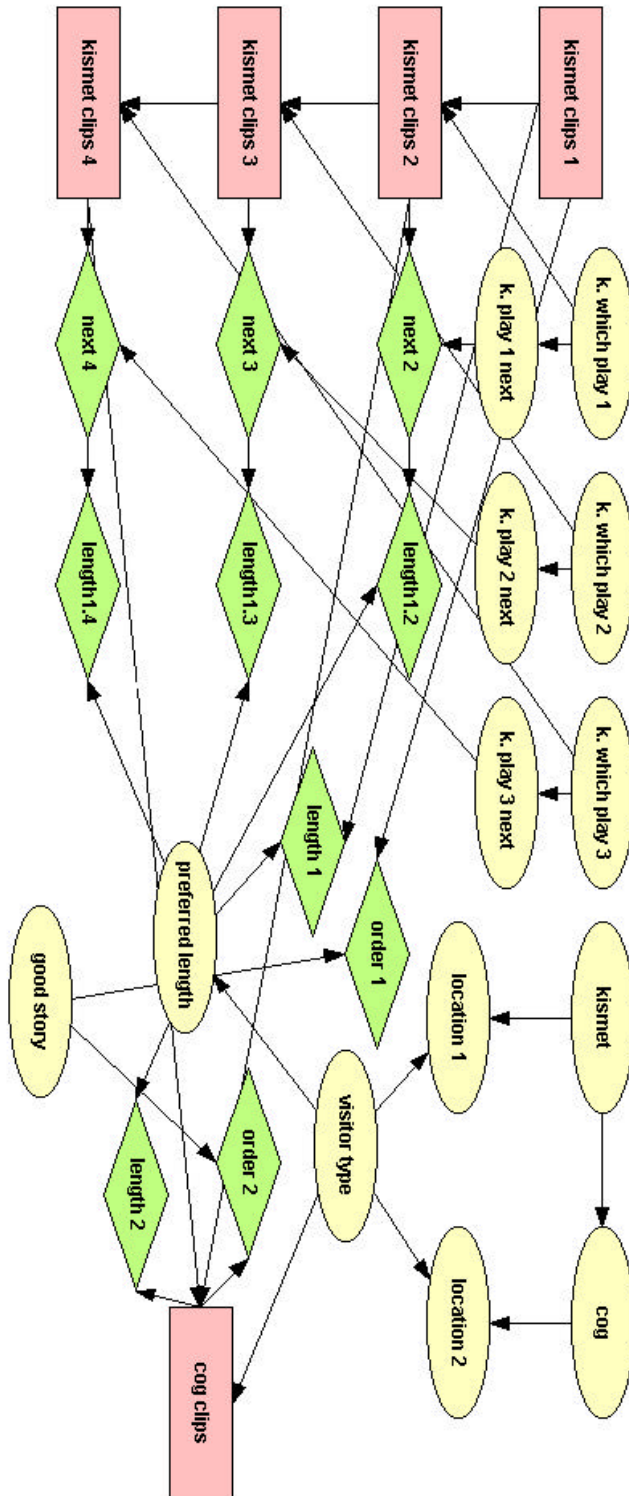


Figure 90. Content editing with with sto(ry)chastics

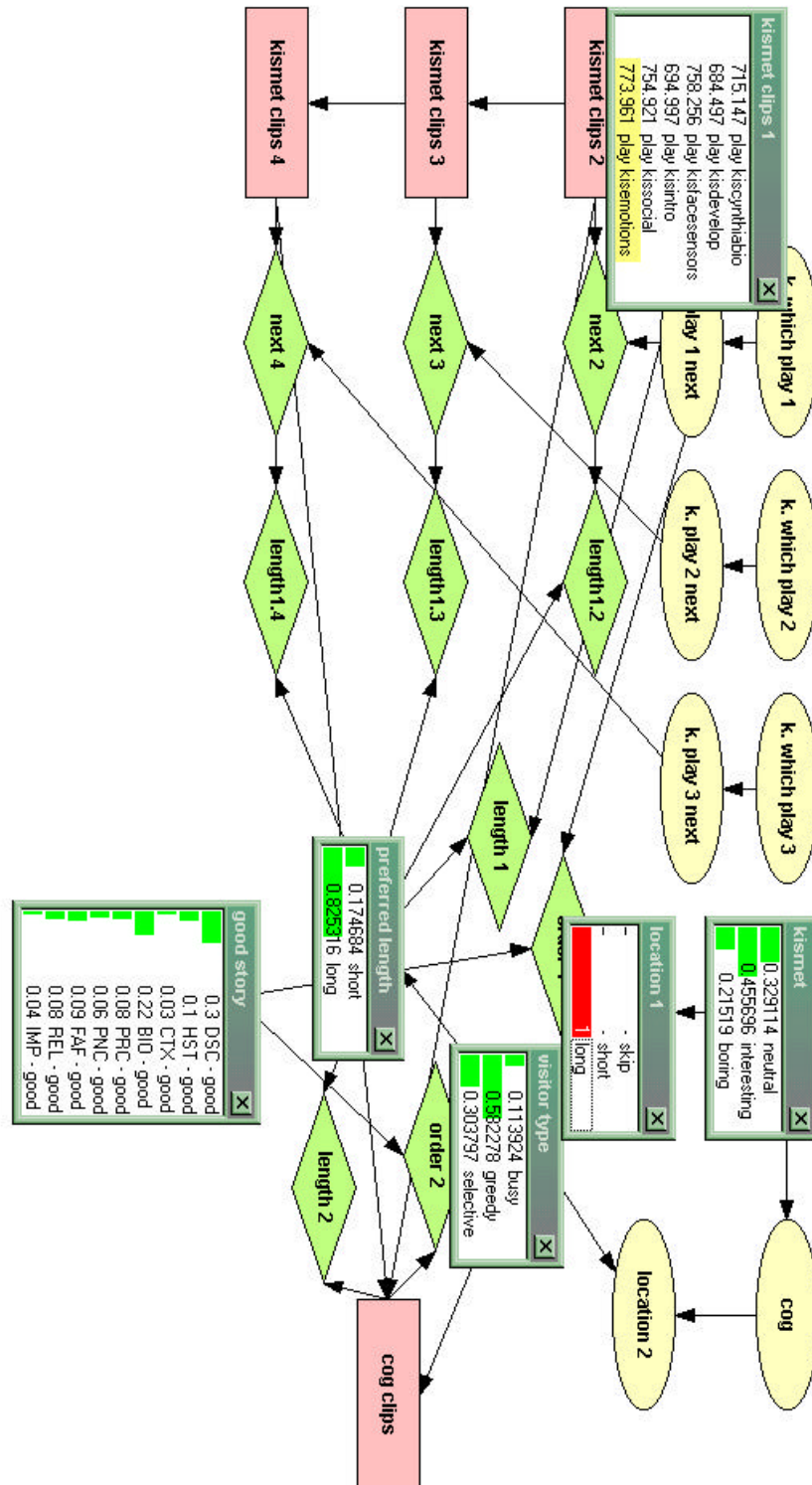


Figure 91. Content editing with with sto(ry)chastics: selection of first segment





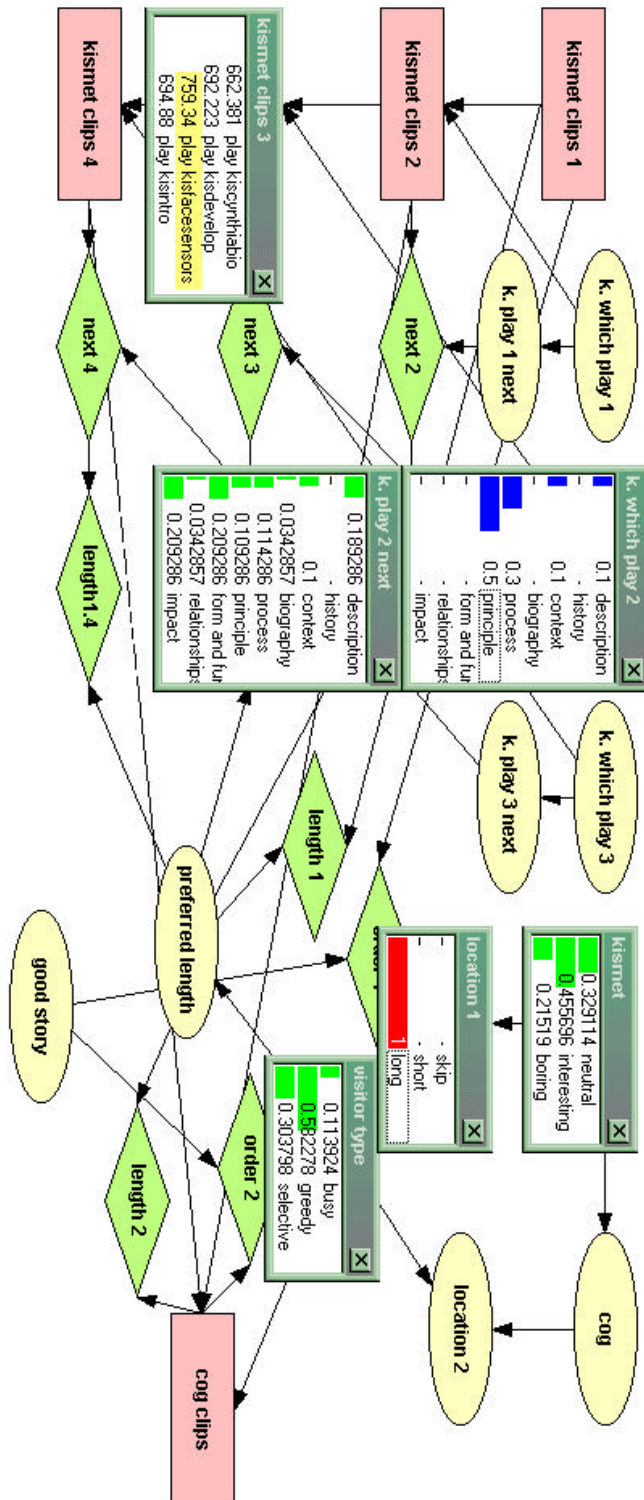


Figure 93. Content editing with sto(ry)chastics: selection of third segment



## 5.3. Content selection for different visitor profiles

The museum wearable prototype, described in detail in Chapter 6, implements visitor type identification and content selection with a Bayesian network as illustrated in Chapter 4 and Sections 5.2. and 5.3. Bayesian networks however allow the system modeller to extend the system much further than what was physically realized in the prototype in the limited time available. I have therefore built and tested various extensions to the Bayesian network presented so far to illustrate extensibility by showing how the system allows the designer to easily add more content, to model more visitor types, and to add more sensors to achieve a more accurate identification of the visitor's interests and to perform robust sensing. All the results of this sections are obtained in simulation, using the Hugin Bayesian network development environment, and should be considered as a first step towards the museum wearable vision described in Sections 3.1. and 3.2. Future work will actually turn these simulations, and the corresponding sensor assembly, into real time software running on the museum wearable.

### 5.3.1. Adding content

It is often the case that after the opening of an exhibit, more content becomes available, and that the curator may wish to add it to the set of audiovisual material of the museum wearable. One of the great advantages of Bayesian networks is their flexibility and the ease given to the modeller in adding, or removing, nodes or states of the system without having to perform any further calculations. As an example, let's assume that three new video clips become available. To add them to the system all that is needed is to update the decision and utility tables of the network. First, the new content segments need to be assigned probability weights which express what they are about in terms of the themes which describe the targeted exhibit [table 33]. The same weights need to be transcribed in the corresponding order *utility* table, multiplied by a balancing factor (300), as before. Finally the utility *length* table is updated with the corresponding length seconds for the new clips. No further calculations are needed. It would be even simpler to remove a segment from the ones available: all it would require would be to simply delete the segment's field from the corresponding decision node, and the system would automatically delete all references to it from the other connected nodes.

CATEGORIES / TITLES		KisNEWFuture	CogNEWBrain	CogNEWSensors
Description	DSC	0.3	0	0.3
History	HST	0	0	0
Context	CTX	0	0	0
Biography	BIO	0	0	0
Process	PRC	0	0.4	0.7
Principle	PNC	0.3	0.6	0
Form & Function	FAF	0.1	0	0
Relationships	REL	0	0	0
Impact	IMP	0.3	0	0
Total P		1	1	1

Table 33. Theme categorization for the new content segments

preferred length	Long	preferred length	Long	
kismet clips	play kisNEWFuture	cog clips	play cogNEWBrain	play cogNEWSensors
	44		90	123
preferred length	Long	preferred length	Short	
kismet clips	play kisNEWFuture	cog clips	play cogNEWBrain	play cogNEWSensors
	-44		-90	-123

Table 34. Lengths in seconds of added clips in for the utility node (up): these are positive values when long segments are preferred, and negative values when short segments are preferred

kismet clips	play kisNEWFuture								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	90					90	30		90
cog clips	play cogNEWBrain								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
					120	180			
cog clips	play cogNEWSensors								
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good
	90				210				

Table 35. Value for the ORDER utility node: these are the same values in table 33 multiplied by a weight of 300

To illustrate how the system works with the new added clips, I include in figures [94,95] the results of probability update after a short stop for the first object (the robot called Kismet) followed by a long stop for the second object (the robot called Cog). The reader can verify that the new segment *kisNEWFuture* competes closely with *kisIntro* to be the preferred clip. Although shorter, *kisNEWFuture* is not the highest ranked because *kisIntro* has more description elements, and description is the favored theme, as indicated in the good story node. The same happens for the second object for which *cogIntro* is still first ranked, followed by *cogNEWSensors*. Note that in the second case, even though the visitor made a long stop, the system still consider that it is a busy type, even if by a short margin. It will require a third long stop for the system to attribute a higher probability to the busy type. If the visitor makes a short stop duration at the second object, instead of a long one, while the system still decides for a busy type, now with a higher probability, the segment selection is different [figure 96]. This is because a different probability for the busy type causes a different probability for the preferred length segment which in turn causes shorter length clips to be preferred in the [short, short] stop duration case, as opposed to the [short, long] stop duration case. This is yet another advantage of using Bayesian networks which handle probability propagation across nodes. A deterministic system, which does simply one to one mappings between inputs and outputs would have not been able to get to the same conclusions for the busy type.

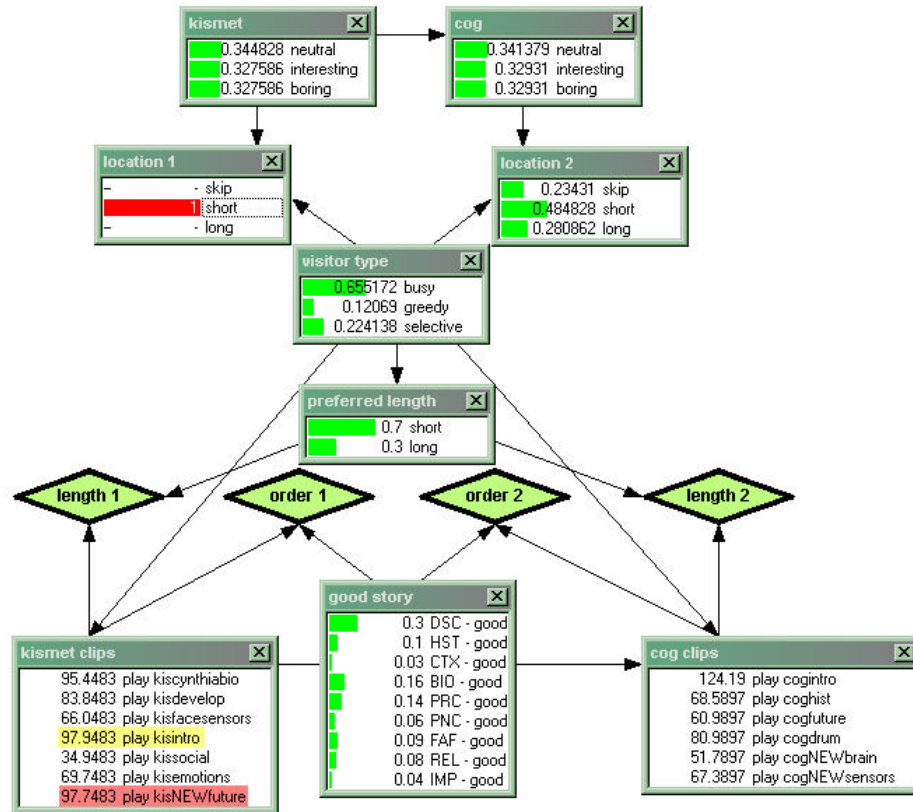


Fig.94

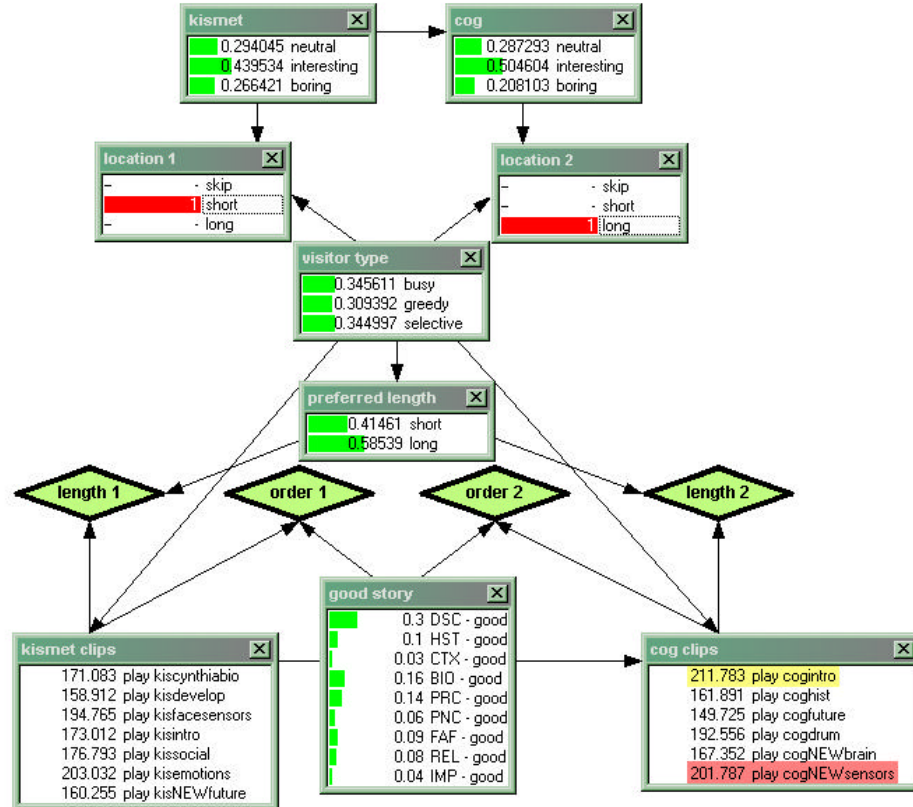


Fig.95

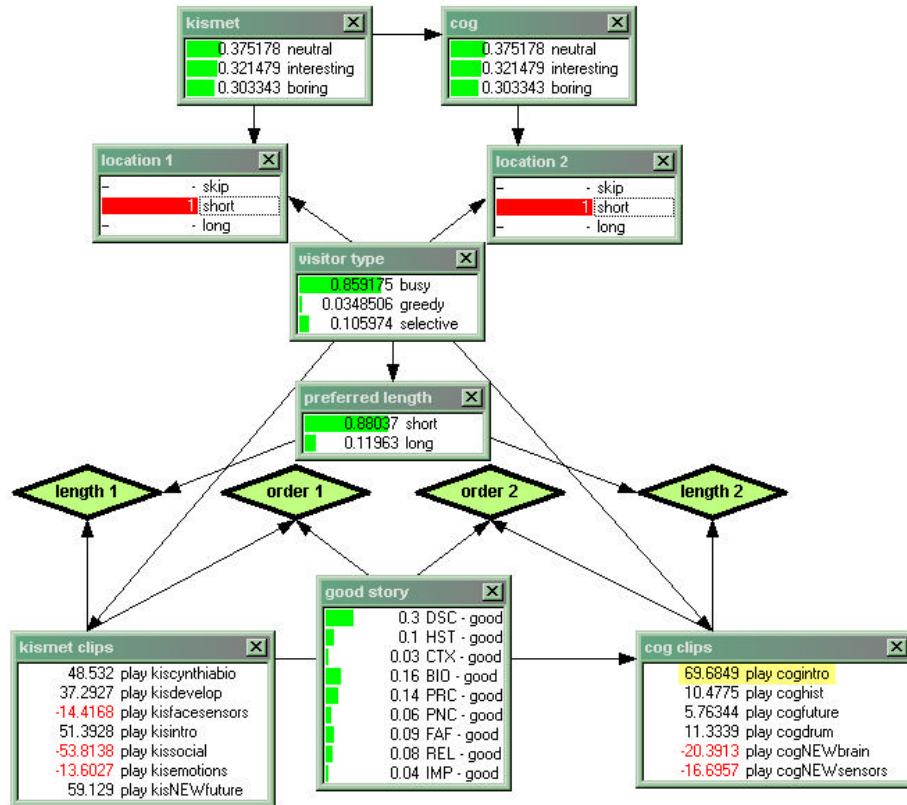


Fig.96

### 5.3.2. Adding visitor types

Adding a new visitor type is just as simple. During a personal interview, Beryl Rosenthal, director of exhibitions at the MIT Museum, described observing a “stroller” type: a visitor who wonders through the exhibit space with no particular strategy or goal. Such visitors spend a random amount of time with each object, and their behavior is somewhat erratic. If they find something of interest they will make a long stop at that object, but, as opposed to the selective type, they may also as well not make other long stops with object closely related to, or belonging to the same theme of, the object which originally attracted their attention. Therefore the conditional probability table for the visitor node, with this new added type, look as [table 36]:

Conditional probability table for the visitor node			
	skip	short	long
Busy	0.2	0.7	0.1
Greedy	0.1	0.1	0.8
Selective	0.4	0.2	0.4
Stroller	0.333	0.333	0.333

Table 36. Conditional Probability table for the visitor type with the new stroller type

The full table, which extends the one showed in section 4.2.4., including conditional probabilities for interesting and boring objects is shown below [table 37]. The remaining tables are updated as shown in section 5.2. The *prior* probability for the type is still equally distributed amongst all the types, but now with four types in place of three it has a different value:  $p(\text{busy})=p(\text{greedy})=p(\text{selective})=p(\text{stroller})=0.25$ .

	neutral				interesting				boring			
	busy	greedy	selective	stroller	Busy	greedy	selective	stroller	Busy	greedy	selective	stroller
skip	0.2	0.1	0.4	0.333	0.1	0.05	0.1	0.167	0.35	0.2	0.65	0.666
short	0.7	0.1	0.2	0.333	0.6	0.05	0.3	0.167	0.6	0.2	0.15	0.167
long	0.1	0.8	0.4	0.333	0.3	0.9	0.6	0.666	0.05	0.6	0.2	0.167

Table 37. Conditional Probability table for the visitor type with the new stroller type with the object type as parent

Note that In the preferred length node the I have given the stroller no real preference about segment length, in accordance to the definition of this type:

Preferred length				
	busy	greedy	selective	Stroller
short	1	0	0.2	0.5
long	0	1	0.8	0.5

Table 38. Preferred length node values

I show in figures 97-101 an example of identification of a stroller type. To do so, I added the nodes corresponding to the third object, as at least three stops are needed to identify the stroller type. This is due to the fact that the stroller will likely do a short or long stop, or will skip, and given that the location node has three states (skip, short, long) three time slices are needed to identify the stroller. The two cases show a visitor making 1. [long, skip, short] stop durations and 2. [skip, short, long]. If the location node has a higher discrete resolution for the time spent by the visitor with the corresponding object, such as the five discrete states: skip, short, medium, long, and very long, then five time slices would be needed. For comparison with the previous case, I have also included the probability distribution on the previous network, with only three visitor types, with the same evidence: 1. [long, skip, short] and 2. [skip, short, long]. In both cases, the three-visitor network identifies a selective type. What this means is that adding a new state for the visitor node allows the system to have higher discriminative power and to be able to distinguish a true selective type [figure 101] from a stroller type.

Note that the *phantappsurg* segment is preferred both in the three types and four types case both for the selective and stroller type. The reason is that is segment has a great utility value on the description theme, and therefore best matches the curator's preferences expressed by the good story node.

Pahntom clips	play phantom-application-surgery									
good story	DSC, good	HST, good	CTX, good	BIO, good	PRC, good	PNC, good	FAF, good	REL, good	IMP, good	
	180				90					30

Table 39. Content theme utility values for the phantom-application-surgery segment

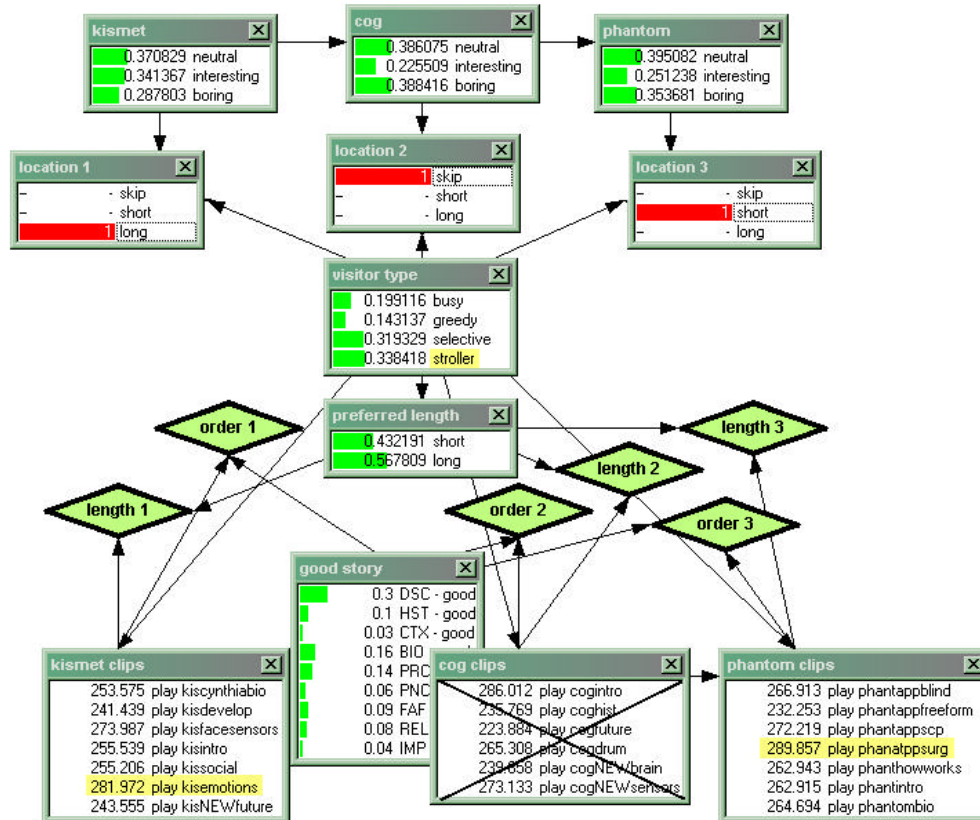


Fig.97

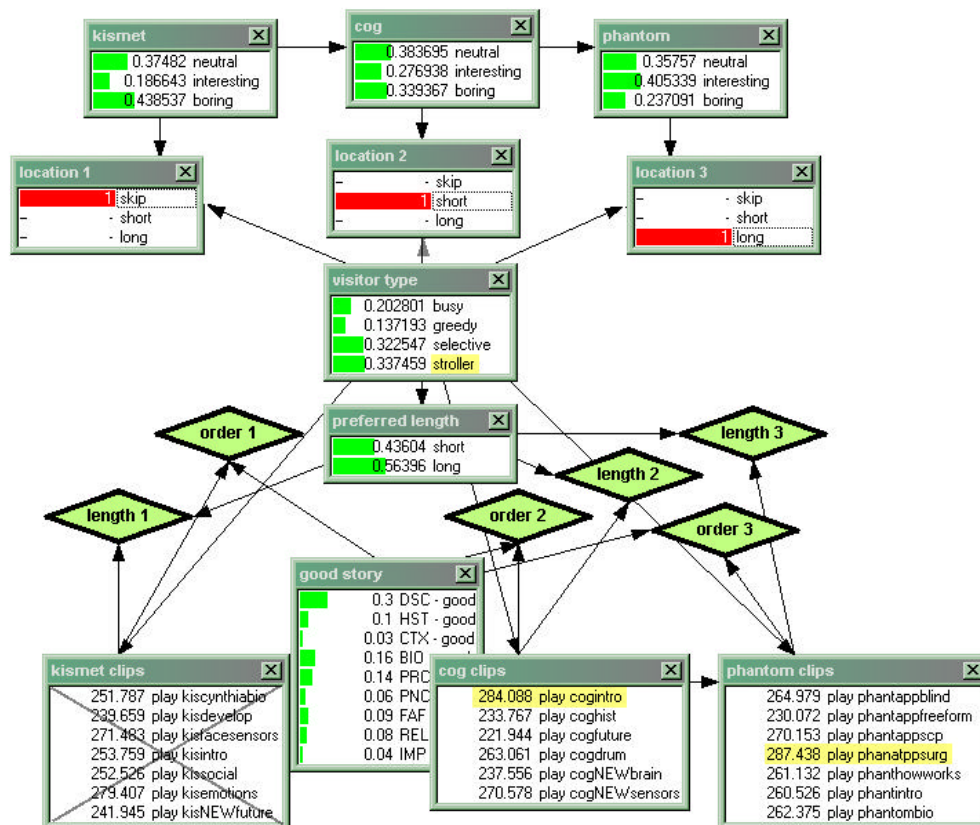


Fig.98



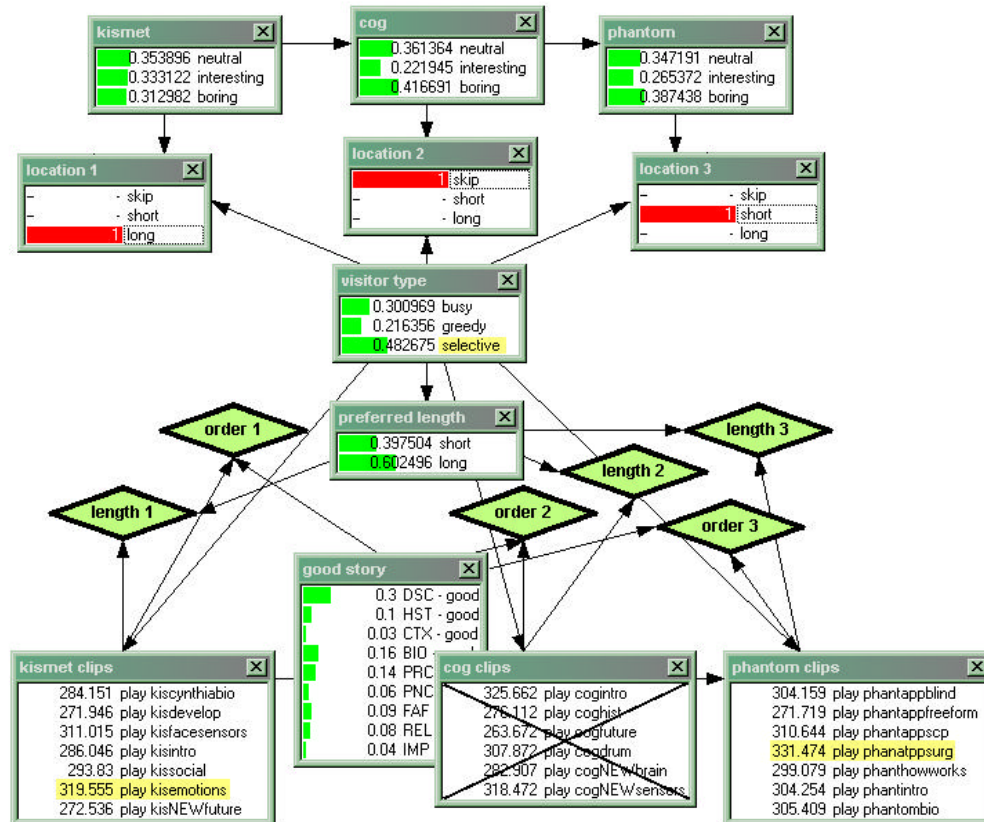


Fig.99

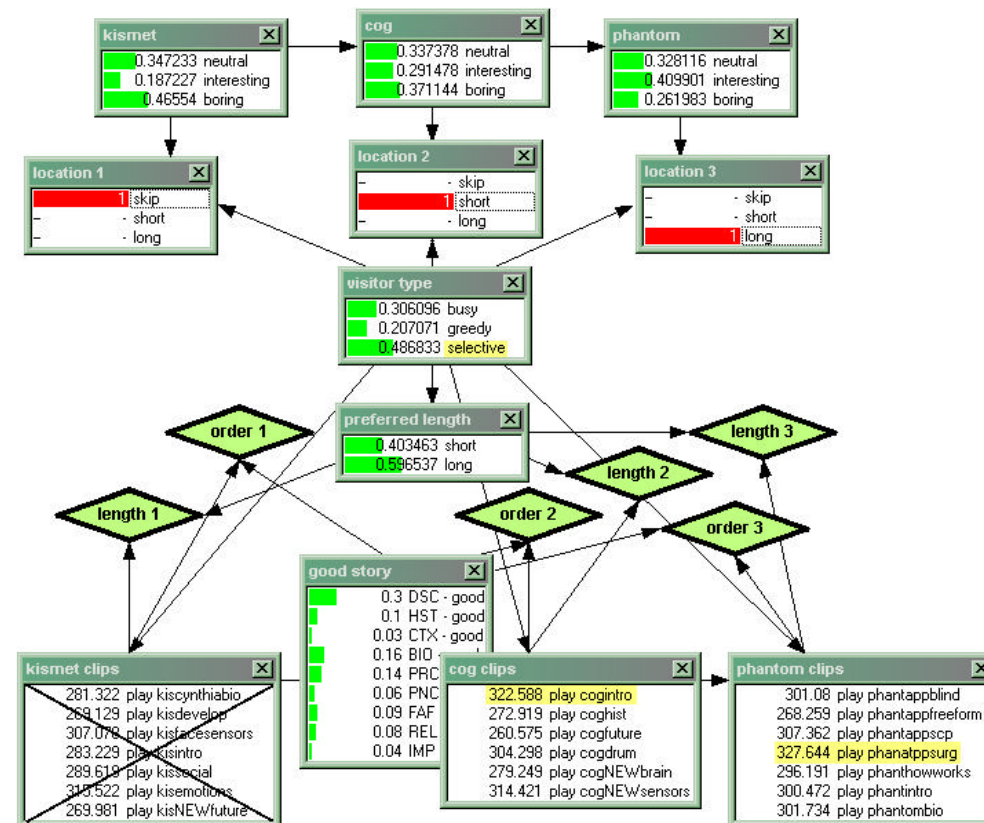


Fig.100

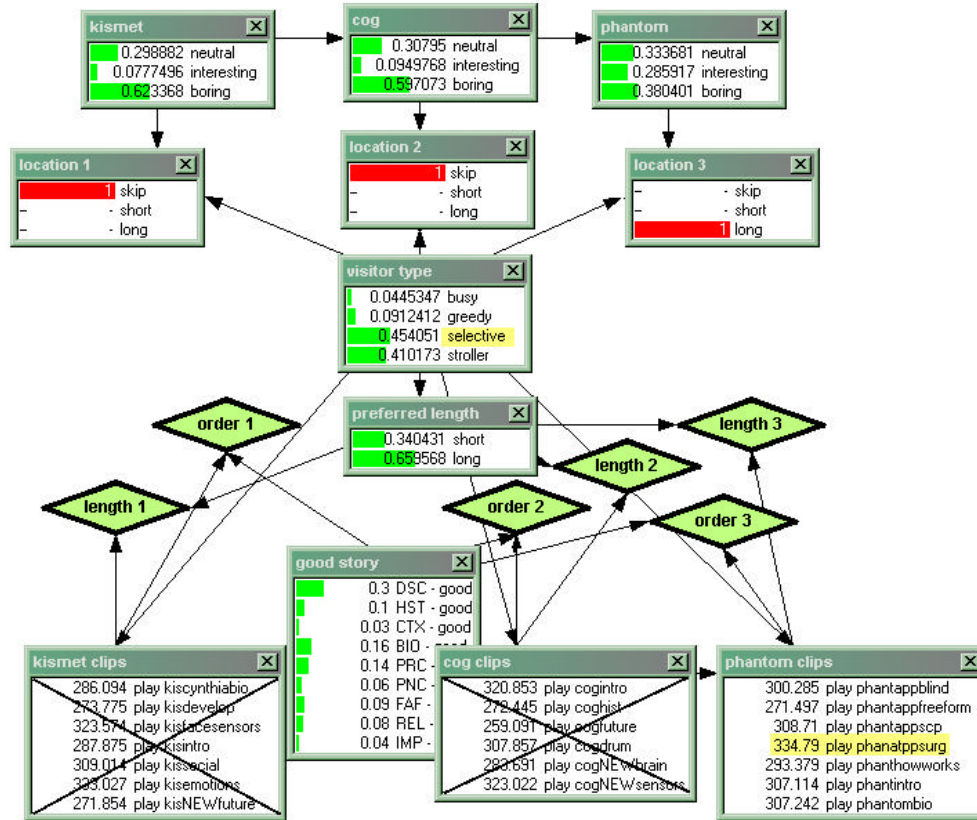


Fig.101

### 5.3.3. Adding sensors: sensor fusion in a simulated environment

The museum wearable prototype described so far relies uniquely on the one location identification sensor. While modeling the sensor measurements probabilistically, rather than in a deterministic way, allows the system to model the measurement uncertainty, it would be still desirable to add more sensors to the wearable so as to have more information about the visitor, or additional robustness with respect to location identification. For content personalization the system should be able to infer an interest profile for the visitors, in addition to their type as they wonder along the exhibit gallery. With respect to the definition of story given in Chapter 5.1., an interest profile in the context of this research means a rating of preference for the story themes given in table 18. Some people for example may like to hear more about the author's biography, and inspiring philosophy of thought, while others may be more interested in the processes and techniques of creation of the artwork. Houbart [Houbart, 1994] uses this type of user profiling, based on story themes, for her Viewpoints on Demand system which edits offline a personalized documentary for the viewer after his/her preferences for the available story themes, and maximum story duration time, are introduced in the system. The museum wearable should be able to perform both visitor profiling and segment assembly according to the visitor's preferences in real time.



Two sensors that could easily be added to the museum wearable to gather this type of additional information are a GSR (galvanic skin response) sensor, and a small camera, placed on the head mounted display, as shown in Chapter 6. While I have performed preliminary studies and signal characterization for these two sensors, it was not possible, for time constraints, to include them as part of the museum wearable physical prototype. What this section presents however is a full simulation on how to extend the Bayesian network presented so far to include these two additional sensors to learn more and provide more to the museum visitor.

### 5.3.3.1. The GSR sensor

The GSR (Galvanic Skin Response) sensor responds to skin conductivity and is often used in the medical and psychological field as an aid to monitor an individual's level of excitement or stress [Healey, 1999]. Psychologists use for example a GSR sensor whose signal is set to be proportional to the velocity of a toy electric train to monitor a child's excitability or response when asked delicate questions. Healey [Healey, 2000] has demonstrated that a series of peaks measured from a GSR sensor, is correlated to a high level of energy or excitement of the human subject. The basic idea behind this sensor is that high energy, excitement, or stress in most individuals, corresponds to increased perspiration, and therefore increased skin conductivity. Two small plates in contact with the skin measure the current that goes through them, which gives a measure of skin conductivity. The main problem with GSR sensors is setting a baseline level for the measurements, as skin conductivity varies across individuals. Most GSR devices require a fairly quick initialization procedure, which lasts about 10 seconds, in which the sensor baseline is set.

A GSR sensor can potentially give very useful information to the museum wearable. First of all it can be easily worn as a wrist bracelet [see Chapter 6]. Second, we could use its measurements, in conjunction with a story segment payout to infer the visitor's interest profile. If for example the GSR sensor measures a train of peaks when the wearable is playing a segment with biographical information about the portrayed artist, the system can infer, *with a certain probability*, that the visitor has a strong interest for this topic i.e. biography. It will then update the visitor interest profile with the gathered visitor preferences. The probabilistic framework offered by the Bayesian network approach is particularly relevant for this type of sensor. For example the sensor could measure excitability for other reasons than that a compelling video segment being shown, such as meeting a friend, or recalling something that happened earlier during the day.

The decision node, in the earlier examples, was weighting segment length, and theme as expressed by the "good story" node, corresponding to the curator's preferences. It now also needs to take into account the visitor's preferences, which may compete with the curator's ordering preferences, to come up with the best content selection for each object. The new extended network, shown in figure 102 models a hypothetical GSR sensor added to the museum wearable and its influence on content selection.

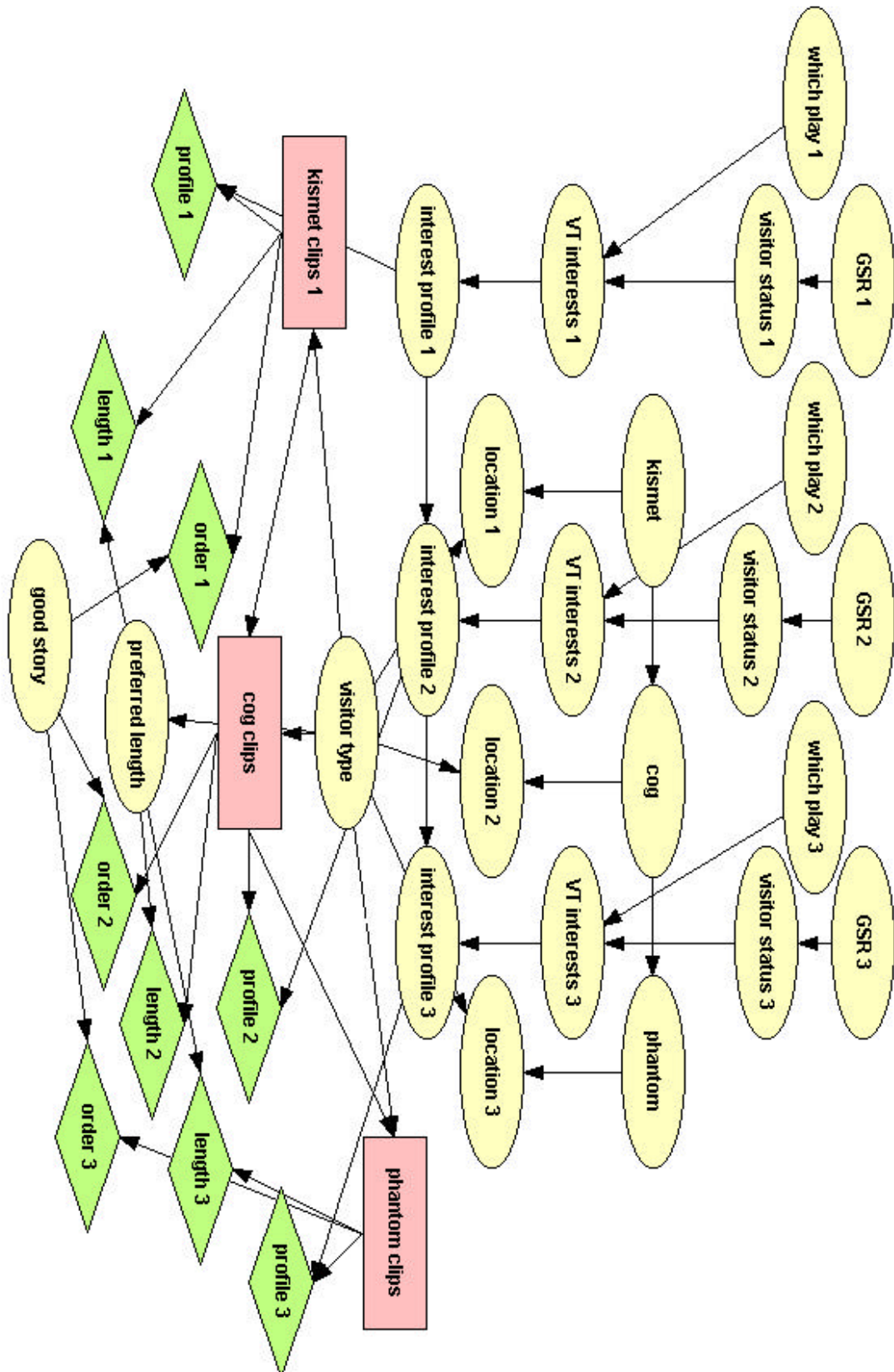


Fig.102. Simulation of visitor's interest profile identification using sto(ry)chastics

Table 40 show the states, prior probabilities, and conditional probabilities for the GSR sensor node and the visitor status node.

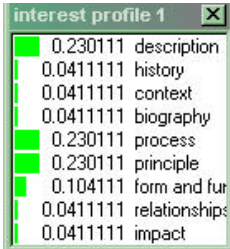
GSR sensor		visitor status	peaks	flat
peaks	0.5	excited	0.9	0.1
flat	0.5	neutral/bored	0.1	0.9

Table 40. States, prior probabilities, and conditional probabilities for the GSR sensor node

Note that as opposed to the visitor node, which is a static node, the visitor interest profile is a dynamic node, i.e. it is repeated for each time slice. The profile utility node has the same weights as the order utility node. This means that the curator's preferences for segment ordering and the visitor's and weighted equally. Other choices are of course possible, according to the modeller's choice, such as having the curator's preferences matter more or viceversa having the visitor's preferences matter more.

To give an example of modeling of the GSR sensor, I show two test cases, one without [figures 103-105] and one with a GSR sensor [figures 106-108]. In both cases the visitor makes a long stop duration at the first object, followed by another long stop at the second object, followed by a short stop at the third object. In the first case, without the GSR sensor, the first segment selection for the three objects is: [*kisemotions-kisfacesensors-kisintro*, *cogintro-cogdrum-coghist-cogfuture*, *phatappsurg*]. In the second case the network uses the information from the GSR sensor to build the visitor's interest profile distribution.

Object 1. As the visitor is seeing the last segment: *kisintro*, the GSR sensor observes a train of peaks. The visitor status node, now consider the visitor to be excited about this clip with 0.9 probability. Therefore the system infers that the system is interested in the topics shown by the *kisintro* clip, and updates the visitor interest profile node [interest profile 1, next].



interest profile 1	
0.230111	description
0.041111	history
0.041111	context
0.041111	biography
0.230111	process
0.230111	principle
0.104111	form and fur
0.041111	relationships
0.041111	impact

Object 2. As the visitor is seeing the last segment: *cogfuture*, the GSR sensor observes a train of peaks. The visitor status node, now consider the visitor to be excited about this clip with 0.9 probability. Therefore the system infers that the system is interested in the topics shown by the *cogfuture* clip, and updated the visitor interest profile node [interest profile 2, next].



interest profile 2	
0.152271	description
0.031311	history
0.031311	context
0.031311	biography
0.152271	process
0.467271	principle
0.071631	form and fur
0.031311	relationships
0.031311	impact

Object 3. Because of the influence on the content selection node by the visitor interest profile, the segment chosen for object 3 is now: *phantombio*, as opposed to *phatappsurg* which was the segment selected in the previous case, based uniquely on the visitor type.

Introducing the GSR sensor, and extending the Bayesian network as shown, allows therefore the experience designer to better tailor content to the user, and to better understand and characterize the user's interests.





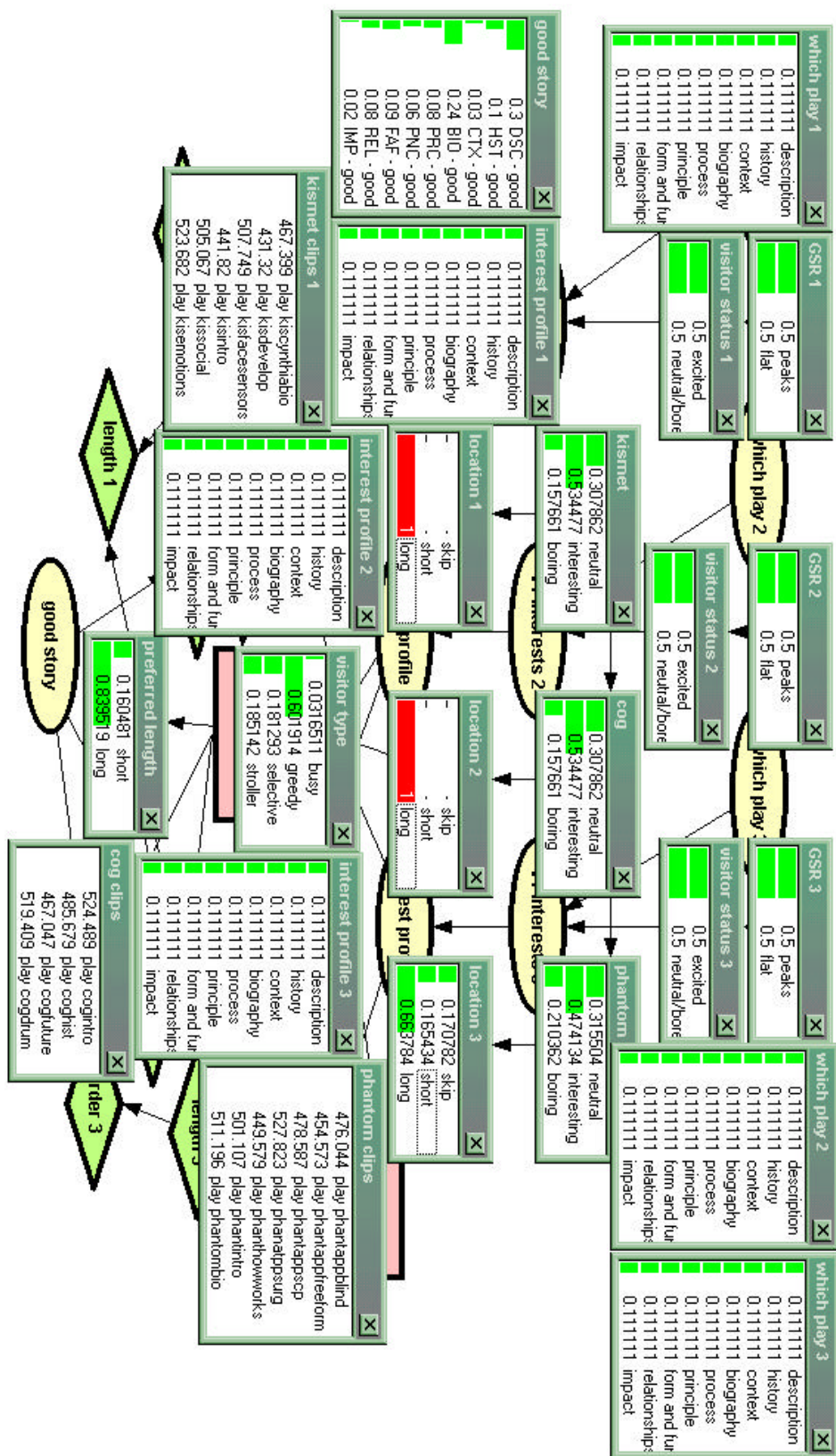


Figure 104. Simulation of sto(ry)chastics without a GSR sensor (for comparison): long stop at the second object.



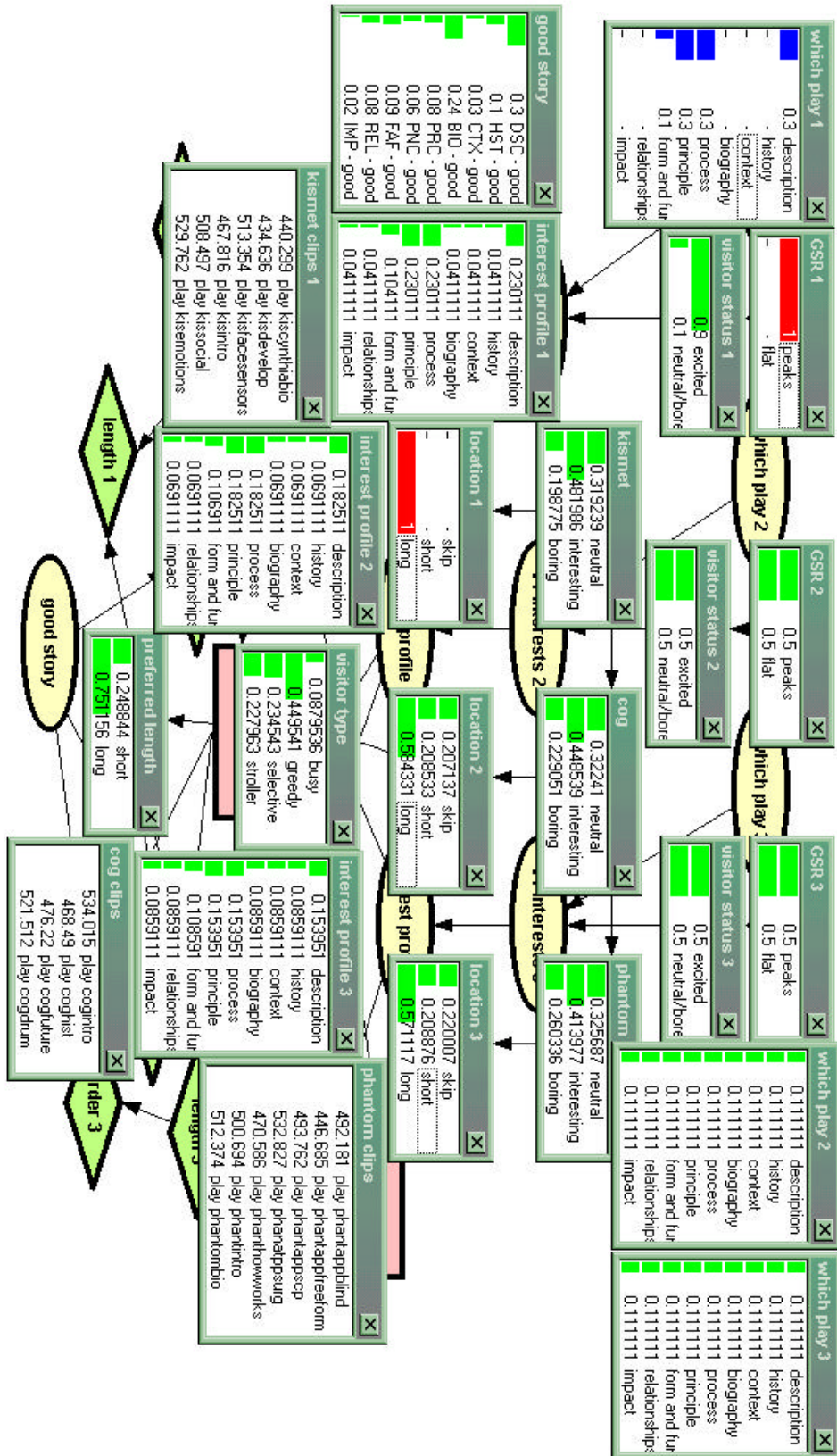


Figure 106. Simulation of sto(ry)chastics with a GSR sensor: long stop at the first object.



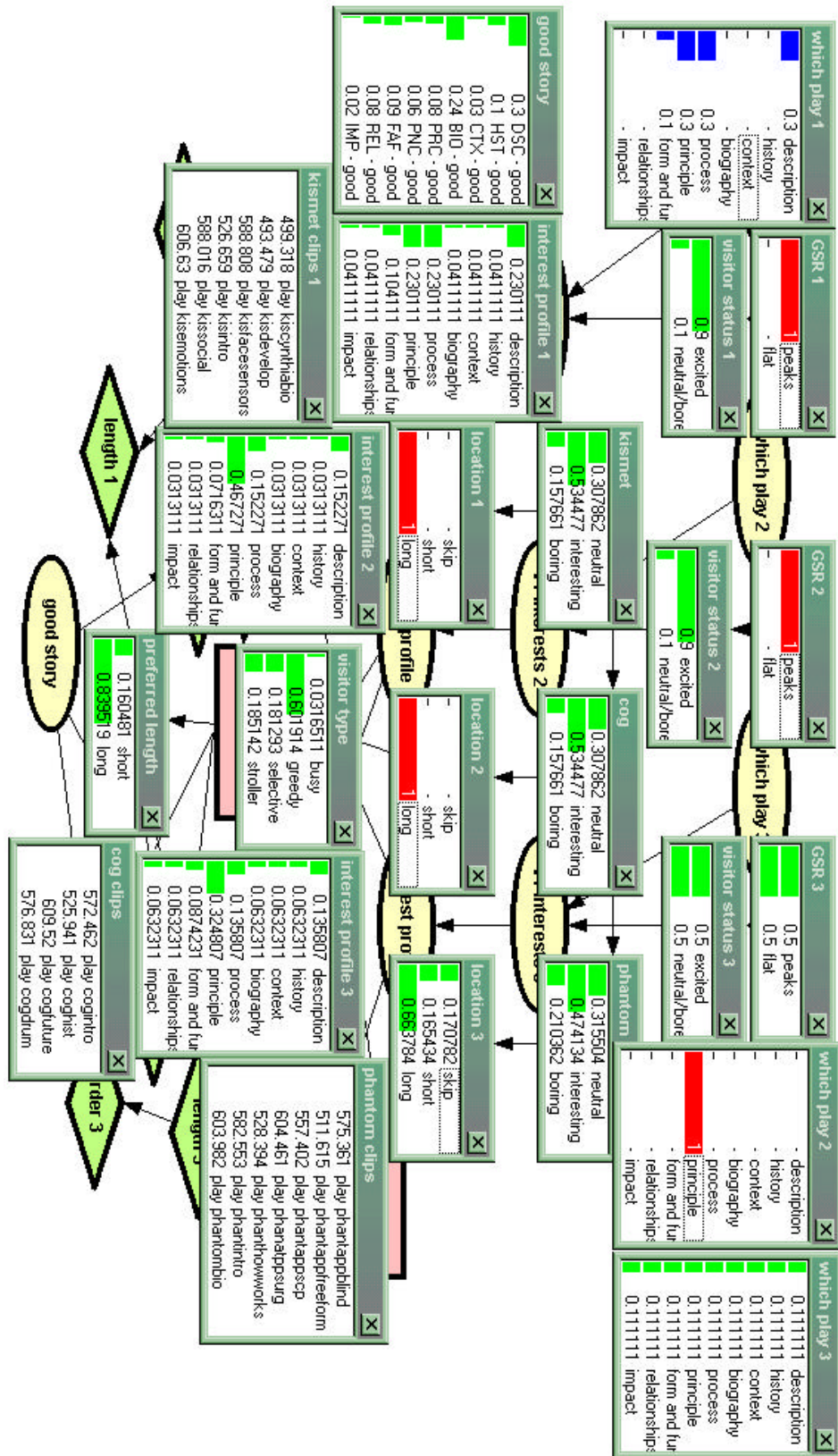


Figure 107. Simulation of sto(ry)chastics with a GSR sensor: long stop at the second object.



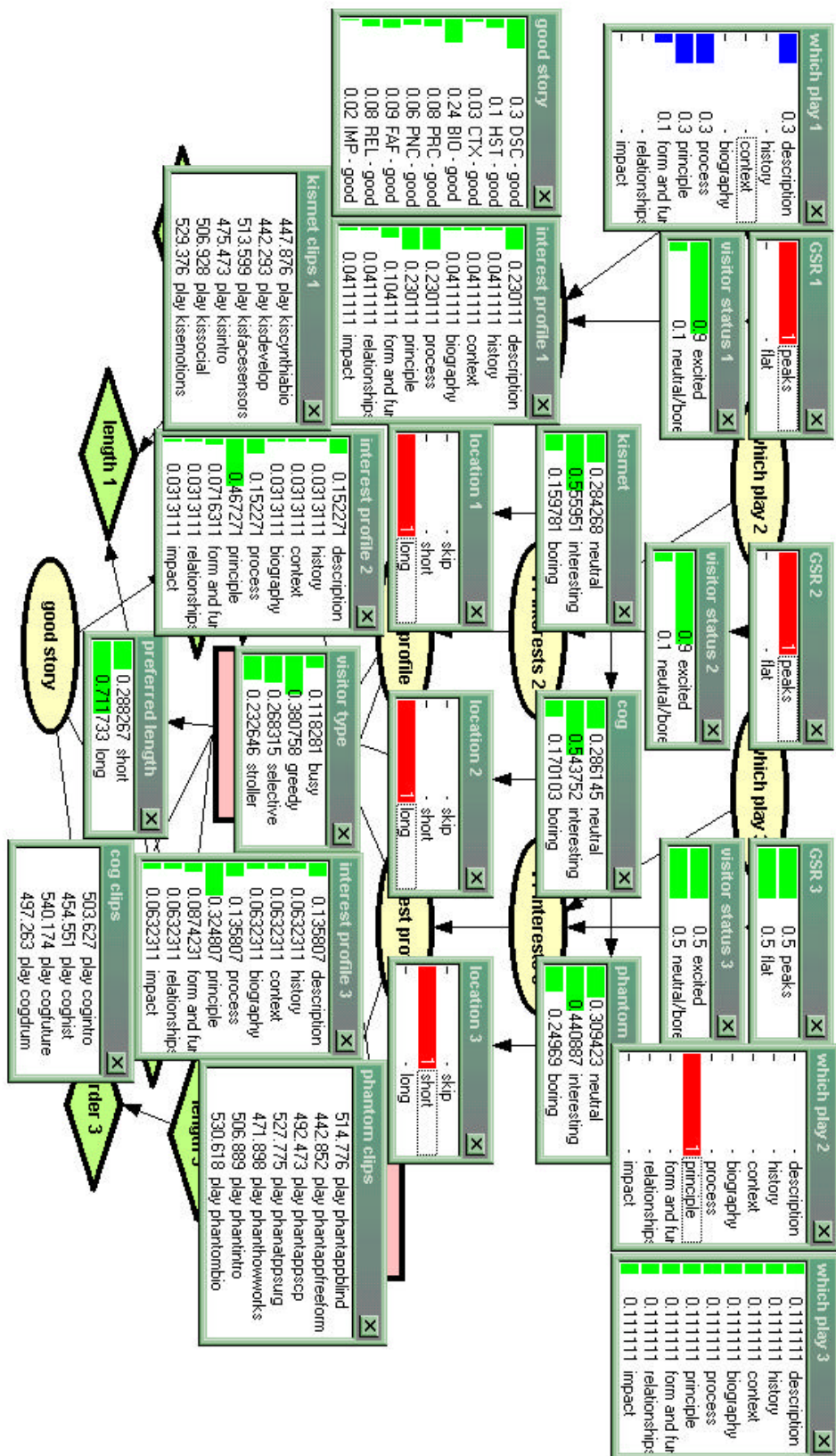


Figure 108. Simulation of sto(ry)chastics with a GSR sensor: short stop at the third object.

### 5.3.3.2. The camera

While the custom location sensor built for this project, described in Chapter 6, has great reliability and long range, relying exclusively on the infrared location sensor to gather information about the visitor's location in the museum and length of state, can occasionally produce errors. The infrared receiver sensor is located on top of the headphones with attached display that the visitor wears, and it can sometimes be ineffective either because it can be covered by hair, or because the headphones may be worn with the headband not vertically aligned. Vertical alignment of the headphones' headband is important so that the tiny infrared sensor located on top of the headband points towards the infrared emitters located in the light rack on the ceiling. The location emitters could also break, as some the electronic components of the emitter tags, such as the infrared emitting diodes could wear out, or power surges could damage the transformer powering the emitters. To achieve robustness for location identification a "redundant" sensor should be added to the system. A camera is the ideal complementary sensor to the infrared location identification emitter-receiver tags. A study on possible placement of tiny cameras to the head mounted display of the museum wearable is described in Chapter 6. The camera is placed on the head mounted display, pointing outwards, and covering the same field of view seen by the wearer: basically the camera sees what the visitor sees.

I have carried out a preliminary study, which was not integrated in research, on the usage of a very small infrared camera which detects the signal from tiny low-power infrared emitters located near the object. This camera set up was preferred to using a color camera doing image recognition for the Robots and Beyond exhibit at the MIT Museum. The reason is that a color camera is best suited in situations in which the museum wearable is "augmenting" a painting exhibit, and therefore the task of the camera sensor would be to recognize flat, two dimensional objects, (the paintings) on a uniformly colored background (the wall on which they are hung). Several computer vision techniques can be used for this task, and they are dependent on how well the technique identifies color under varying lighting conditions, such as ones determined by a large window in the exhibit area. Recognizing the three dimensional objects – the robots on display – from different viewpoints is a much harder task, and while various approaches in computer vision are also available, further studies would have to be performed to test the reliability of the technique in an fairly unconstrained environment such as the museum gallery. Using object tagging with small, low powered, infrared devices, or even better, with passive infrared reflective material placed right next to the object, provides the means to do location identification with a small infrared camera. The simulation presented in this section, is based on the latter type of infrared camera/infrared object tagging system.

The camera based location identification sensor tells us more than the infrared location identification sensor. While the latter is only able to say if the visitor is in proximity of the tagged object, the camera sensor would be able to detect when the visitor is actually looking at the object. This is a typical example of sensor fusion used to achieve sensing robustness (see Chapter 2): the two sensors have some overlap, but they also give complementary information in other respects. If the camera can be powered

with the wearable's battery (it absorbs about 2W) without wearing battery duration, and if the tags can be placed near the museum objects in a way which is aesthetically acceptable by the exhibit designer, both the camera and the infrared location sensors should be used. From the camera it may also be possible to gather information on the visitor's level of attention: if the wearer is fixating an object, it is possible to gather, with a certain probability, that they are attentive and interested. If instead the visitor is in proximity of an object, and yet looking around, that may be a sign of distraction or non interest. Finally, if the system receives a location identification from the camera but not from the IR sensor, it might reasonably deduce that there is something wrong either with the IR receiver on the wearable or the emitters.

This situation is modeled by the network in figure 109, which extends the story selection Bayesian network to include additional modeling for the camera and the IR sensor, separate from the time spent at each location, which is encapsulated in the "location" nodes. To allow enough room on the page for all nodes in the diagram, only the nodes for the first object have been added. The reader should consider that when used with the museum wearable, the additional nodes to the left for object 1, should also be replicated for all other objects modeled by the network. The initial probabilities and conditional probabilities for the added nodes are given by the tables 41-44. The visitor node has an added state, called "don't know" used to model incertitude about the type when neither sensor is active. Note that from the conditional probabilities assigned to the "visitor present at location 1" node, the system considers the IR sensor more reliable than the camera sensor, as it assigns a higher probability of presence of visitor given IR signal detection (0.9) than given the camera based object detection (0.8). These same probabilities are usually higher, i.e. 0.98 and 0.95. These values have been set to highlight the effect of these two sensors can have on the visitor type identification, as explored by the example below.

Camera 1		IR signal received	
Looking at object 1	0.5	IR present	0.5
Not looking at object 1	0.5	IR not present	0.5

	Camera 1	
visitor status at 1	Looking at 1	Not looking at 1
attentive	0.8	0.4
distracted	0.4	0.6

Visitor present at location 1				
Camera 1	Looking at object 1		Not looking at object 1	
IR signal received	IR present	IR not present	IR present	IR not present
VT at 1	1	0.8	0.9	0.02
Not present	0	0.2	0.1	0.98

Tables 41,42,43. Probability tables for the nodes of the Bayesian network which simulates a museum wearable with a camera sensor

IR problem				
IR signal received	VT at 1		Not present	
Visitor present at location 1	IR present	IR not present	IR present	IR not present
IR malfunction	1	0	0	1
IR OK	0	1	1	0

Table 44. Conditional probability table for the IR problem node

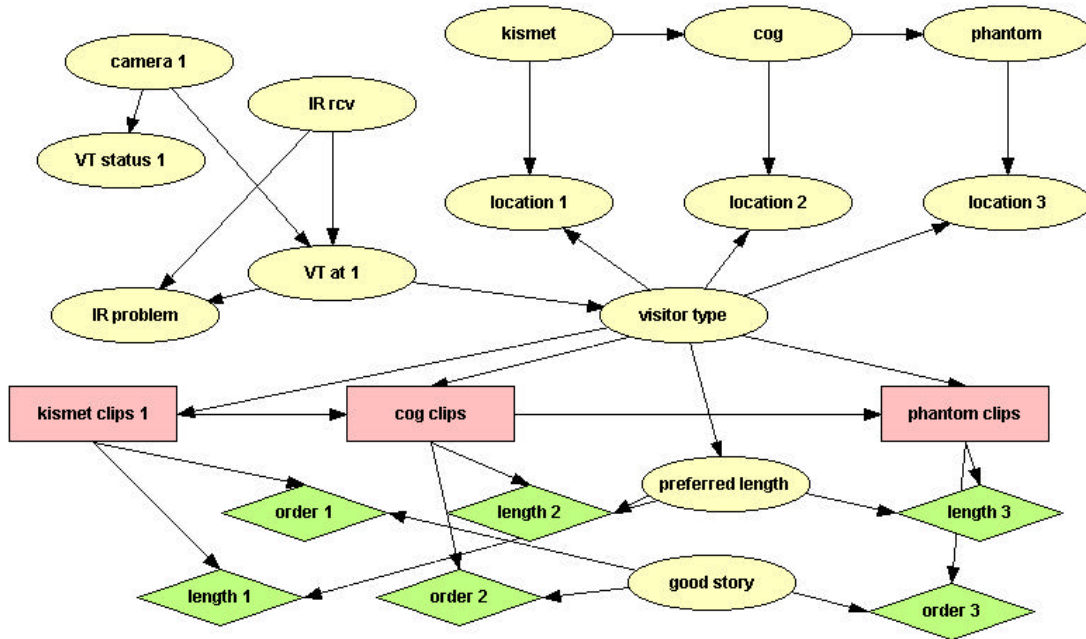


Figure 109. Simulation of sto(ry)chastics with a camera.

To show the effects of camera modeling, figures 110-113 show the probability distributions on this network, in the four following cases:

1. The IR sensor detects the visitor's presence, but the visitor is not looking at the object . The visitor makes a short stop at the first object.  $P(\text{VT at 1})=0.925$  and  $p(\text{busy}|\text{short})=0.595$ .
2. The camera detects that the visitor is looking at the first object, but the infrared location sensor does not. For a short stop,  $p(\text{VT at 1})=0.846$ , less than in the previous case, because the camera sensor is considered less reliable than the infrared location sensor. This has an effect on the type estimation as now  $p(\text{busy}|\text{short})=0.551$ , less than in the previous case. Also note than in this case the network signals a high probability of infrared sensor malfunctioning, as expected.
3. Both the camera and the infrared location sensor detect the visitor's presence and the visitor makes a short stop. Therefore  $p(\text{VT at 1}) = 1$  (the system is sure

because both sensors agree) and  $p(\text{busy}|\text{short}) = 0.636$ , higher than in the previous cases.

4. Neither sensor detects the visitor's presence. In this case the network correctly says that it does not know what type is the visitor as the posterior probability for the don't know state for the type node is: 0.89, and  $p(\text{visitor}=\text{busy})=p(\text{visitor}=\text{greedy})=p(\text{visitor}=\text{selective})=0.0358$ .

The camera also gives information on the visitor status. If the visitor stares at an object for a long time, we can reasonably think that he/she is attentive, as opposed to distracted. This can have an impact on the visitor's preferences as shown in table 45, containing numbers extracted from the simulations in figures 114, 115. When the visitor is attentive, we can model the network for attentiveness to have an impact on the visitor's profile, whose values are increased. As shown in these examples, attentiveness causes higher values to appear in the visitor profile node, and as a consequence a different ranking is given to the content segments for the next object.





Distracted visitor	Attentive visitor
	
	

Table 45. Different values for the visitor profile in case of an attentive vs a distracted visitor.

While the previous case with a camera and a IR sensor provided an example of sensor fusion at the feature level [Hall and Llinas, 1997] this last case provides an example of sensor fusion at the strategy level. In the previous case the system was trying to establish how certain is the presence of the visitor at a certain location (feature) and both sensor contribute to determine that value. In this case, the sensors cooperate in determining a more accurate user model or profile.



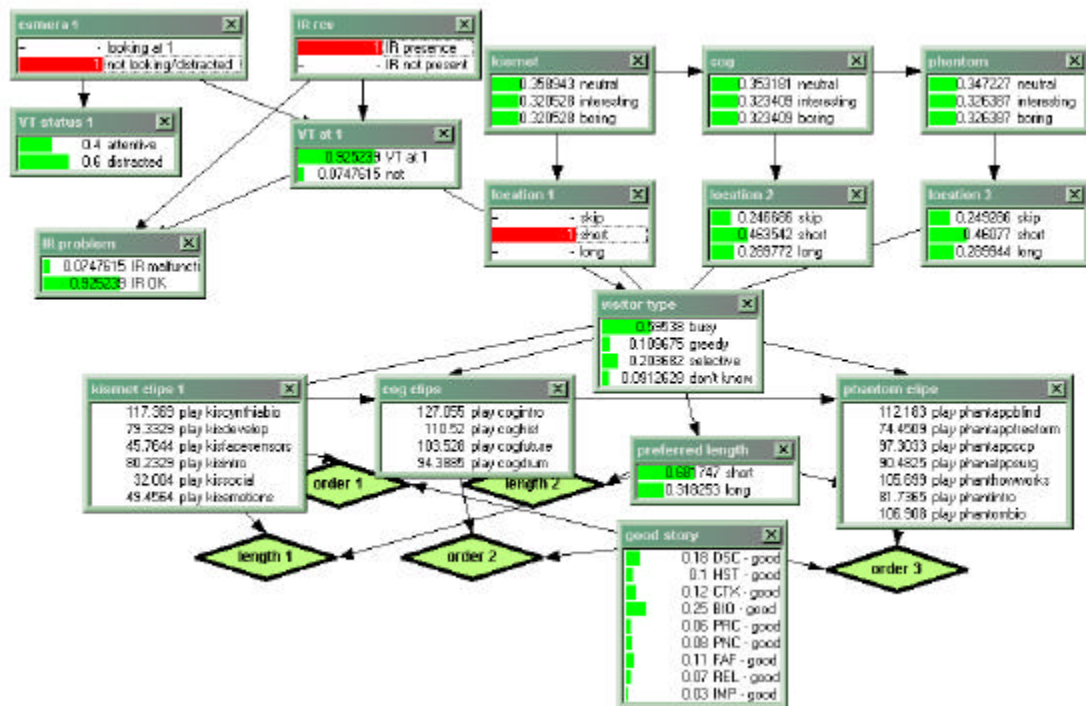


Figure 110. IR sensor detects the visitor, visitor distracted.

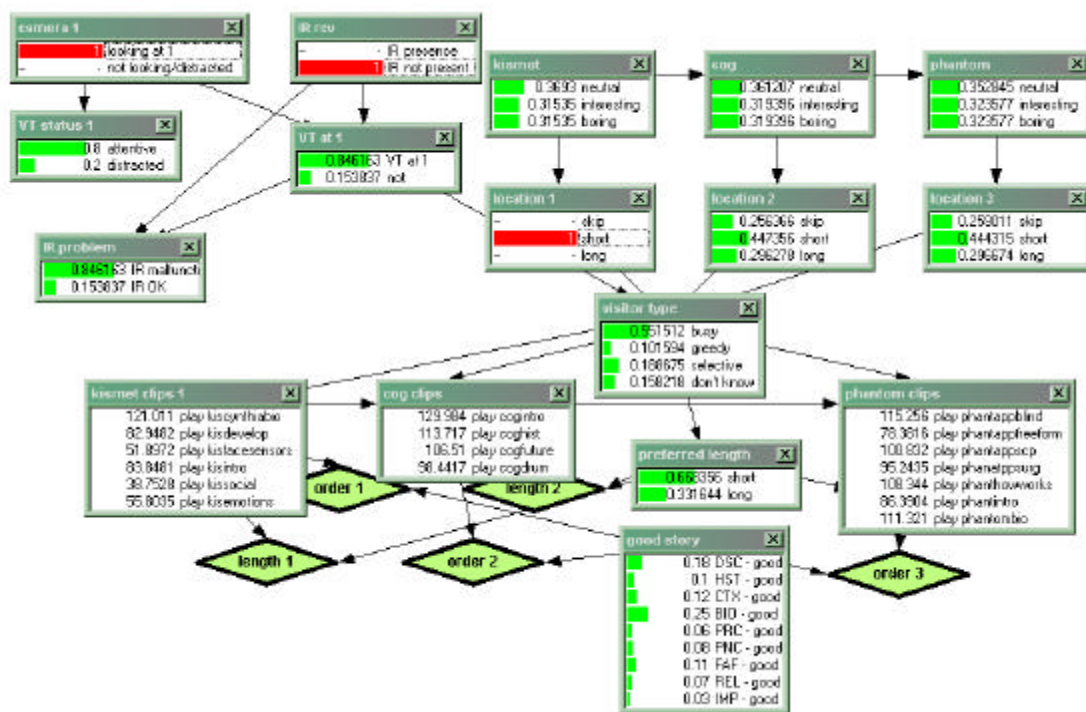


Figure 111. Camera detects the visitor is looking at the first object, but IR makes no detection.

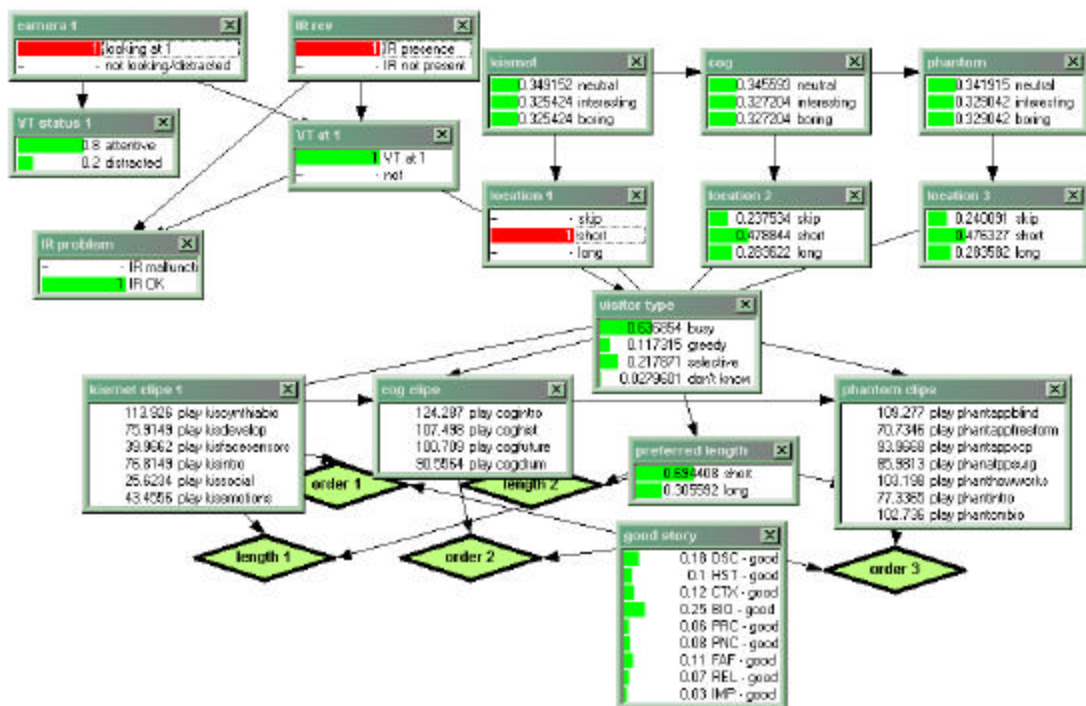


Figure 112. Both the camera and the infrared location sensor detect the visitor's presence.

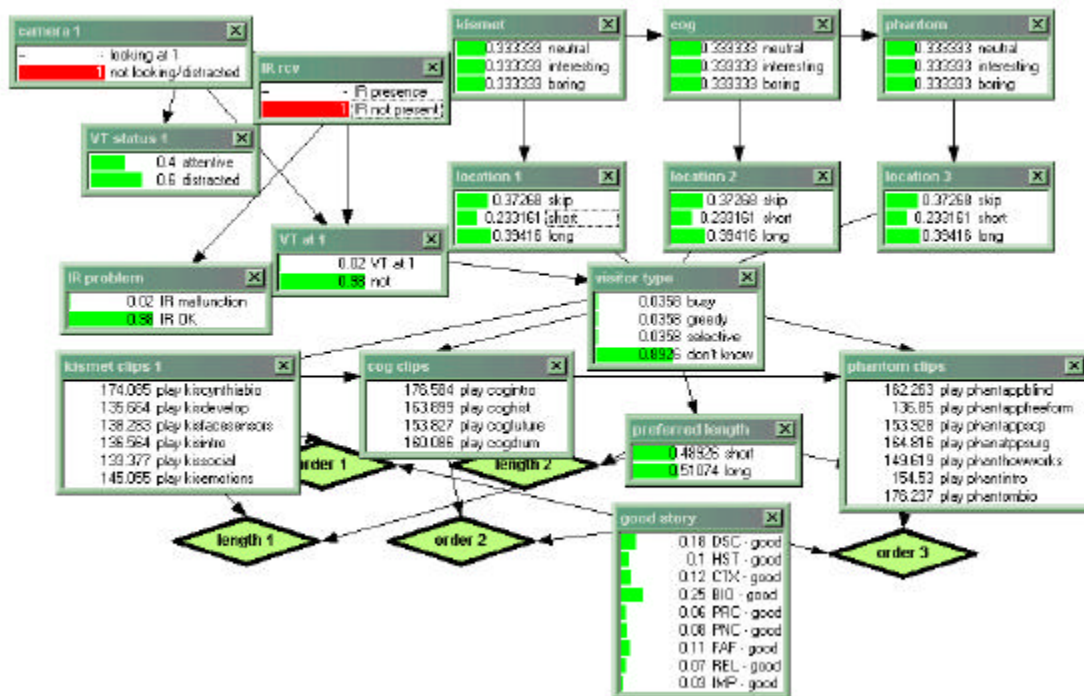
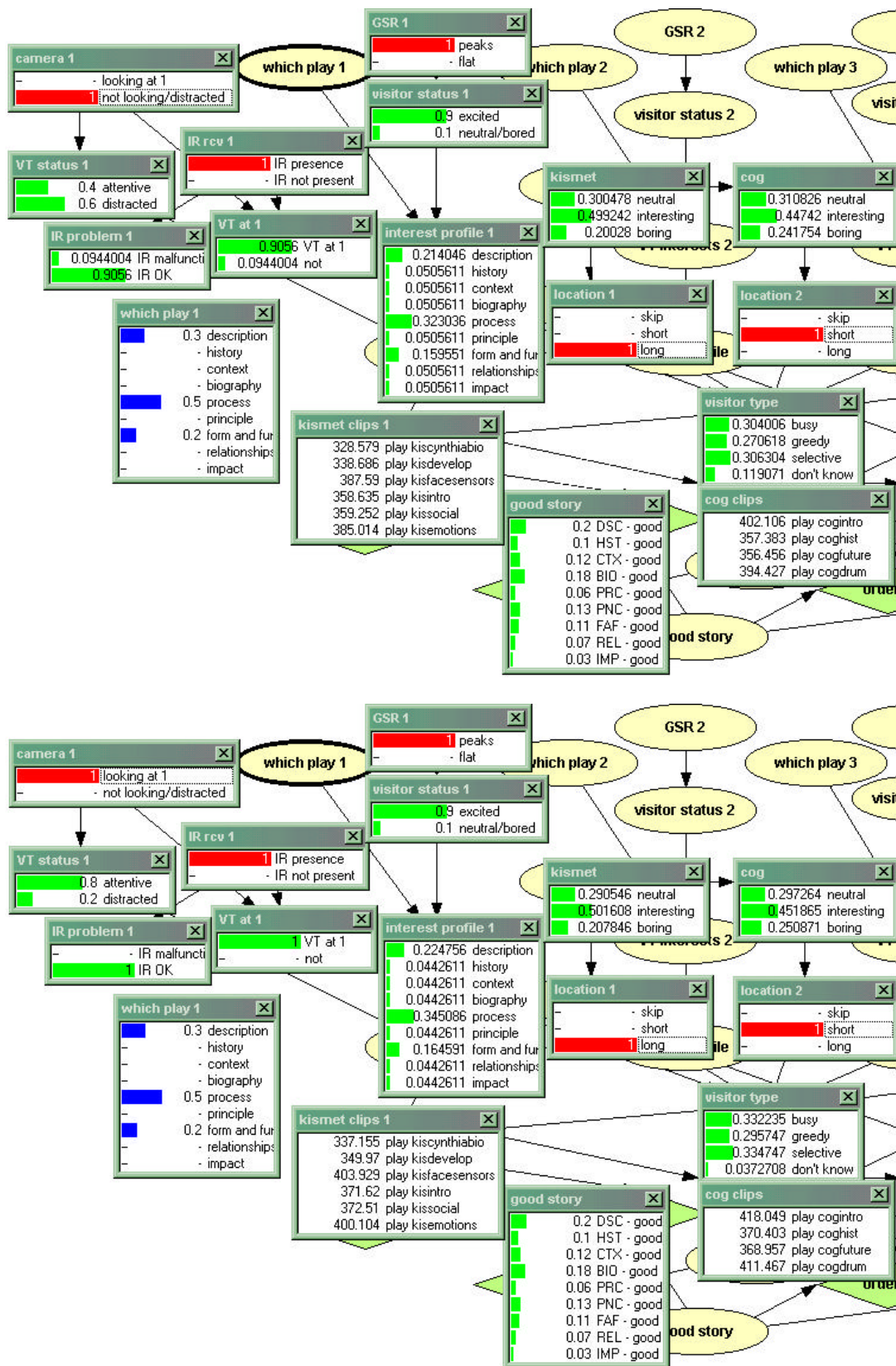


Figure 113. Neither sensor detects the visitor's presence.





Figures 114 and 155. Visitor interest profile and ranking for segment selection if visitor distracted/attentive

## Chapter 6

# Building the Wearable

### 6.1. The wearable computer

The museum wearable is made by a lightweight CPU hosted inside a small shoulder pack and a small head mounted display. The display is a commercial lightweight monocular, VGA-resolution, color, clip-on screen, which is attached to a pair of sturdy headphones. When wearing the display, after a few seconds of adaptation, the user's brain assembles the real world's image seen by the unencumbered eye with the display's image seen by the other eye, into a fused augmented reality image [figures 116, 117].



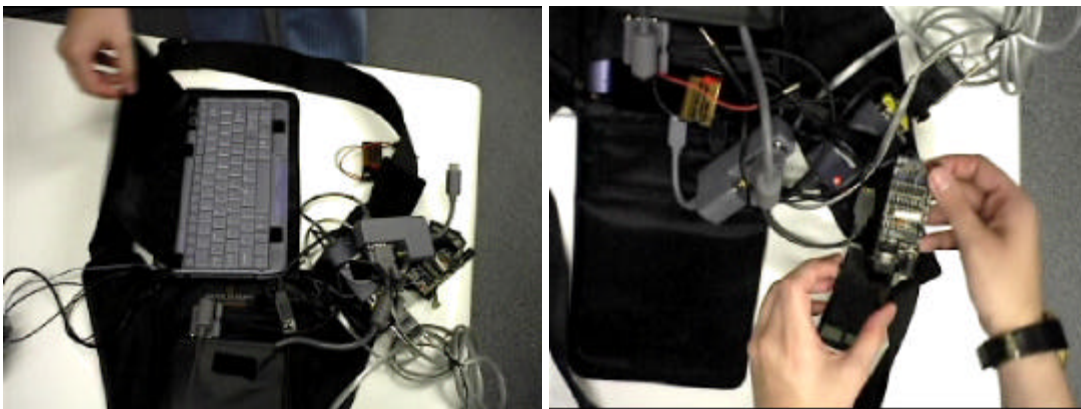
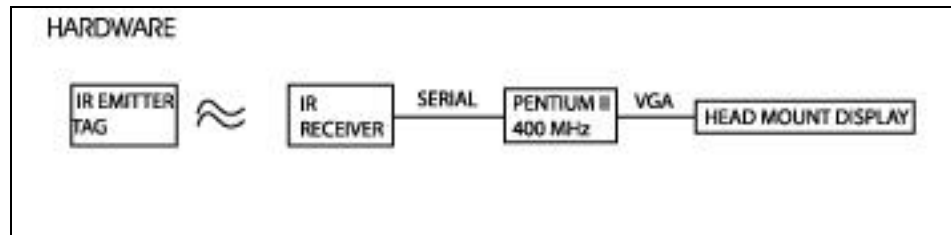
Figures 116, 117. Camera “wearing” the head mounted display: shows how the user’s brain assembles the real world’s image seen by the unencumbered eye with the display’s image seen by the other eye, into a fused augmented reality image.

To monitor the visitor’s behavior in the museum, and deliver a story as a function of the visitor’s evolving path, I have built a first prototype of the museum wearable using uniquely a location sensor. The location sensor informs the wearable on the wearer’s location in the exhibit, and proximity to an object on display, and length of stay for each tagged location. The location system is made by a network of small infrared devices, which transmit a location identification code to the receiver worn by the user and attached to the display glasses. The transmitters have the size of a 9V battery, and are placed inside the museum, next to the regular museum lights. They are built around a PIC microcontroller and their signal can be detected as far as about 30 feet away within a cone range of approximately ten to thirty degrees. The location sensor receiver is made of

two parts: a tiny infrared detector, located on top of the headphones, and the circuitry which detects its signal and transmits it to the wearable computer via the serial port, which is hosted inside the carrying shoulder pack.

Summary of elements of the museum wearable hardware [figures 118, 199]:

- containing shoulder pack
- computer (CPU): SONY picturebook from which the display has been removed to reduce weight
- Head Mounted Display (HMD): VGA resolution MicroOptical clip-on mounted on sturdy headphones with a custom mount
- HMD's powering unit: hosted inside the containing shoulder pack
- Infrared receiver: the sensor is located on top of the headphones and the receiver circuit is located inside the containing shoulder pack.



Figures 118, 119. Hardware parts of the museum wearable. Left: CPU, connectivity, carrying shoulder pack. Right: closeup of infrared location receiver circuit.

To obtain more, and more accurate, information about the visitor's behavior inside the museum gallery additional sensors are desirable. A discussion on which sensors can be added to the first museum wearable prototype, and for which purpose, can be found in the next paragraph. In view of having a museum wearable which can later be expanded to include other sensors, and process information not just from the infrared location sensor, but for example also from a small camera processing images in real time, I have chosen to use and modify a commercially available small sized laptop computer. I selected the SONY picturebook PCG-C1VPK for its combined size, weight, computing power, multimedia capabilities, and longlasting batteries. Given that the images generated by the laptop are viewed uniquely through the head mounted display, I have removed the LCD screen from the picturebook, to reduce weight and size, as shown in figures 118, 121. The picturebook features a Crusoe™ processor TM5600 clocked at 667 MHz, and without the LCD weighs only approximately one lb, and has a size of 0.5" X 9.8" X 6.0" (H x W x D). The picturebook has a 15GB capacity hard drive, which allows the programmer to store on the local hard drive many hours (8-10) of MPEG-compressed VGA resolution video (640x480) (approximately one hour of MPEG-compressed 640x480 video per one GB of available space on the internal hard drive). It also has 128 MB SDRAM, which allows the computer to play smoothly the audio and video clips, as well as process images in real time when the computer is connected to a camera. This computer has outstanding multimedia capabilities: it has an ATI RAGE™ MOBILITY graphics chip with 8.0 MB SDRAM, hardware to encode and decode MPEG1 and MPEG2 digital video, and hardware MIDI for sound. The external ports include one USB port which is connected to the infrared receiver with a USB to serial converter, and a VGA and headphone output which are connected to the video/audio inputs of head mounted display of the museum wearable. It also supports one type II card, which can be used to host a PCMCIA card for wireless communication over the internet or a PCMCIA image acquisition card. All these features, combined with a battery life of 2.5-5.5 hours with the standard lightweight battery, largely enough for a single museum visit, make the picturebook an ideal choice for the selected application.

An alternative to the picturebook is the smaller handheld IPAQ pocket PC 3670. The iPAQ 3670 features 64 MB of SDRAM and a 206-MHz Intel StrongARM SA-1110 32-bit RISC Processor. It has USB or serial connectivity that would interface with the infrared receiver of the museum wearable and it is only 5.11" x 3.28" x 0.62" (HxWxD) in size. To be used for our application it would need a dual PC expansion slot to host a VGA PCMCIA output card, to send the images to the head mounted display, as well as a PCMCIA wireless connection card to allow a video server to stream the MPEG video to the iPAQ computer, as it would not have enough internal storage for all the necessary content.

In order to build a first prototype of the museum wearable, so that it can be expanded to use other sensors, such as a camera for real time image processing, I chose the SONY picturebook. Although slightly bigger (larger) in size, the picturebook has MPEG hardware, as well as all necessary multimedia capabilities and storage space, which have allowed me to focus on the mathematical modeling work described in this document (sto(ry)chastics) rather than in solving hardware problems with a less powerful device,



such as the iPAQ 3670. However, now that the prototype has been built and tested, the iPAQ 3670 is a desirable solution for the deployment of several museum wearables – which need to work only with the location sensor – in a museum. The iPAQ solution is cost effective because its cost is considerably smaller than the SONY picture book, and its size is also smaller.



Figure 120. PAQ pocket PC 3670

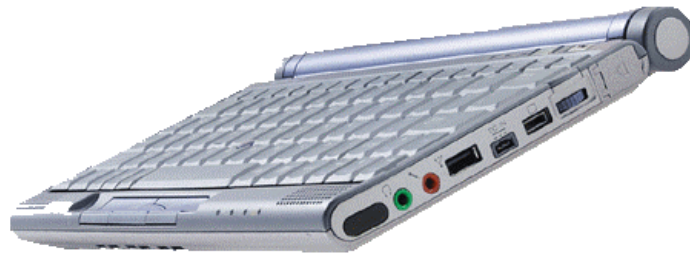
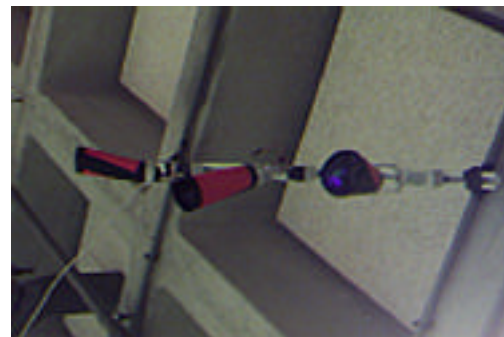
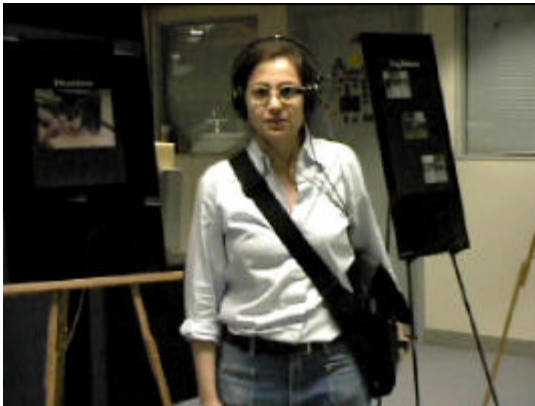


Figure 121. Sony picturebook with removed LCD



Figures 122, 123, 124, 125. Above: visitor wearing the museum wearable and receiving an audiovisual story about the displayed artwork (picture in picture). Below: wearable museum prototype laboratory testing setup.

## 6.2. The choice of sensors

Adding other sensors to the museum wearable would allow the system to obtain more reliable information on the visitor's behavior in the galleries. I performed a study on available off the shelf sensors that, in addition to the above mentioned infrared location identification sensor, can contribute to modeling and understanding the visitor's type and interest profile. The following list identifies these sensors in order of importance:

- IR location sensor
  - Identifies visitor's location and path along the museum. Helps identifying objects or themes of interest by measuring how long the visitor stays in proximity of an object. Helps identifying the visitor's type by determining the if visitor's exploration strategy is made by many short stops (busy type), a few long stops (selective type) or many long stops (greedy type).
- GSR (Galvanic Skin Response) sensor
  - Measures visitor's levels of excitement in response to the content shown on the head mounted display. A high level of excitement suggest, with some probability, that the content category being shown is of interest to the visitor and it is therefore likely to be selected again by the system.
- Computer vision (with IR object tagging)
  - Measures the visitor's level of attention by determining which object the visitor is looking at and for how long the visitor actually fixates the object.
- Motion sensor (accelerometer)
  - Contributes to determine the exact path of the visitor along the exhibit as well as visitor-specific motion patterns, which become meaningful in association with other sensors i.e. approaching an object in a particular way when interested, or suddenly stopping and or coming back when surprised, or moving around erratically when distracted or disoriented.

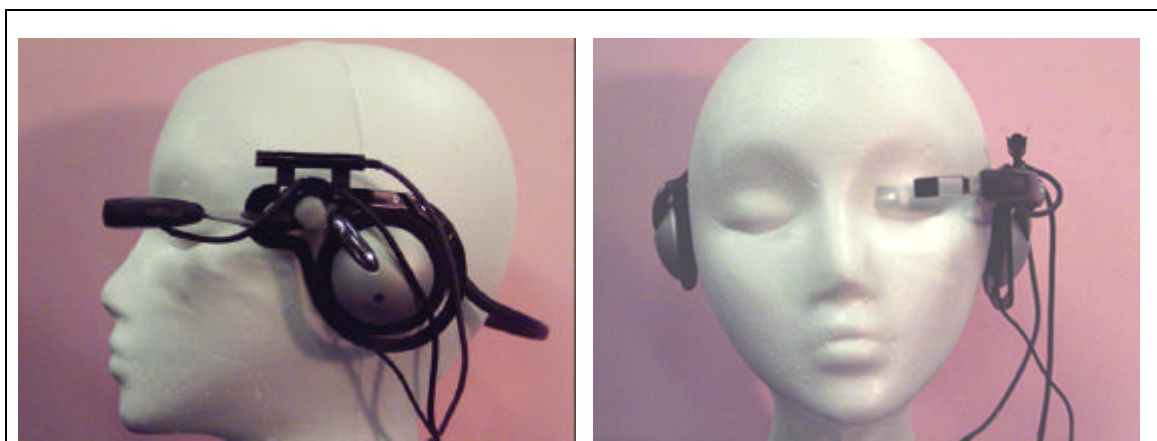
While using only one sensor, may seem like a limiting factor to construct an interactive experience such as the one described, having such long range infrared location ID can provide a great deal of useful information for the targeted application. With the location ID receiver, connected to the wearable through the serial port, we can measure a sampled path of the visitor throughout the exhibit, how long the visitor stays in proximity of the tagged object on display, and his/her overall strategy of exploration. Skipping objects all pertaining to the same category, is an index of dislike for that category, or similarly stationing for a long time next to the legged robot for example, may mean interest for humanoid-like robots.

While the project described in this document features only the infrared location identification sensor, having the GSR sensor would allow the system to identify not only the visitor's type but also the visitor's interest profile, by measuring the level of

excitement of the wearer in conjunction with the content presented on the head mounted display. If for example the visitor reacts with excitement to biographical video clips the system will infer that the visitor likes biography and will use that information to build an interest profile for the visitor. The GSR sensor can be worn as a bracelet around the wrist, as shown in figures 126,127,128. A simulation of how the system would use the GSR sensor in conjunction with the other aforementioned sensors to gather more information about the visitor and match more accurately the visitor's desires and interests is provided in section 5.3.3.



**Figures 126, 127, 128. Study of placement for GSR sensor**



**Figures 129, 130. Study of camera placement on the head mounted display**



## 6.3. Sensor Design: the infrared location sensor

### 6.3.1. Requirements and Alternatives

When I started this project I conducted an extensive search on commercially available off the shelf tagging technology that I could use right away to bootstrap the project. The requirements for a location sensor to be used in a museum, for the targeted application are:

- **functionality:** the location sensor is made of two parts: an emitter, situated in a convenient location in the museum, in proximity of the object that it tags; and a receiver, carried by the visitor, together with the museum wearable.
- **size and shape:** both emitter and receiver need to be of a relatively small size: the receiver needs to be lightweight and small to be easily carried by the visitor. The emitter needs to have a sufficiently small size that it does not stand out or disturb the layout or landscape of the museum galleries. It needs to have a shape or enclosure that are appropriate to the room where it is placed.
- **range:** while range is certainly a function of the museum gallery layout, it is safe to require that the location emitter covers an area of visibility around a museum object, which has the shape of an semi-circle and a radius varying from 2-10 feet away from the object. If the location emitter tags are placed on the ceiling, together with the standard museum lighting, where they would easily be hidden, the requirement for range increases up to 25 feet, for high ceiling placement.
- **directionality** and no overlap: while the range needs to provide a clean signal up to 25 feet, there cannot be overlap between areas that different location identification sensors cover. Therefore the location sensor need to feature a high directionality which can eventually be adjusted around the object of interest.
- **immediacy** of signaling: for the location sensor to achieve its purpose, the receiver needs to receive an identification signal from the emitter as soon as the visitor enters the area covered by the emitter, the sensed area around the object on display. This implies a frequency of signaling of at least 2Hz to allow the system to perform at least minimal error checking.
- **power consumption:** it is important for the receiver unit carried by the visitor not to require much power so that it needs large and heavy batteries, or daily battery changes. Ideally the receiver should be powered by the wearable computer and should not draw more than an absolute maximum of 1 Watt to avoid wearing the batteries that power the wearable computer. Powering the emitter may be more problematic, as an interactive museum exhibit, visited with the museum wearable, needs as many emitters as the are objects on display. If the emitters are battery powered, this imposes the constraint that the batteries last for as long as the exhibit, as a daily change of batteries for all emitters for all objects on display would be highly impractical. Yet to provide

a location signal which covers the range specified above the power consumption would likely be more than what small sized batteries would cover for the whole duration of the exhibit. Therefore powering the location emitters with transformers connected to the available standard power lines is highly advised.

- **connectivity:** the receiver needs to be able to communicate to the wearable its readings from the emitter tags, via the available serial or USB connection.

Keeping in mind the above specs, I conducted a literature and commercial availability search for the desired location identification sensors and examined various possible choices. Following are considerations on some of the solutions I evaluated:

- **passive RF ID tags**  
these involve passive (non powered) small emitter tags that would be placed near the object on display at the museum. The wearer carries a small antenna that capacitively charges the emitter which then sends an identification signal to the receiver worn by the wearer.
  - advantages: ease of placement and availability of emitter tags which do not need to be powered.
  - disadvantages: small PCMCIA receivers were announced not yet available as of January 2001. The available receivers are too heavy and power hungry to be used in a wearable application. The available range with the prototype not commercially available PCMCIA receivers is only about 7 feet, barely adequate for the museum wearable.
  - comments: it would be an ideal solution if it had an adequate range and small sized receivers were available. Probably will be in a year or two as hardware developers make new progress.
- **Infrared tags, based on the IRDA protocol**  
[<http://www.media.mit.edu/~ayb/irx/irx2/>].
  - advantages: they are small, lightweight, easy to build and customize.
  - disadvantages: the range is too short, only up to 7-10 feet in low light conditions. They are also too directional as they cover only a cone of emission of infrared light of a very narrow angle: they can be sensed only along a line.
  - comments: they would be a good solution, if the emitters can be powered via a transformer from the standard power line, and if they can emit a stronger and better shaped signal.
- **Image or object recognition by real time computer vision**
  - advantages: it does not require any physical tagging of the objects on display and therefore it is easy to set up and allows the exhibit designer to easily change object positions and reconfigure the exhibit layout, if needed. It does require training the recognition parameters of the program on the existing objects to be recognized.

- disadvantages: while it is possible to recognize paintings on a wall if they are sparsely located and the wall has a uniform color, the image classification is dependent on the changing light conditions of the museum rooms. This adds an extra variable in the image classification process which makes the recognition less reliable than in any of the previously mentioned cases. Recognizing three dimensional objects is also much more complicated than recognizing paintings on a wall, and therefore using image recognition is not a general purpose tagging option.
  - comments: image recognition is an interesting solution in those limited cases in which the images to recognize can be easily segmented from their background (i.e. large paintings on a white wall) and where it is possible to reliably model the varying light conditions in the galleries. It is however not reliable enough unless further effort is invested in more robust classification techniques.
- Bluetooth location tags:
    - advantages: many people carry already bluetooth technology endowed objects, such as their cellular phone or PDA. These objects could also receive additional information from the museum wearable, which would allow the visitor to have a continued museum experience also outside the museum walls. The museum wearable would have a personalized exhibit catalogue downloaded to the visitor's PDA, or the next generation cellular video phone could be the wearable computer itself.
    - disadvantages: customization is difficult as prototyping kits are not easily available. Potentially the receiver can be connected to the wearable computer via USB or PCMCIA yet I have not been able to access sufficient hardware or software development resources to customize Motorola's PCMCIA bluetooth receivers nor to program Ericson's bluetooth modules as emitters. Power requirements for the emitter tags are also an issue, as bluetooth is in general a power hungry technology, unless the tags can be powered from the standard power line.
    - comments: the technology is simply not mature yet to build or purchase bluetooth based locations sensors.

Based on the above considerations I therefore realized I needed to build my own location sensor to turn the museum wearable into immediate tangible reality. Following from the previous analysis, I concluded that the solution which included most of the desired features of the location sensor was to improve on the existing and easily customizable infrared tags.

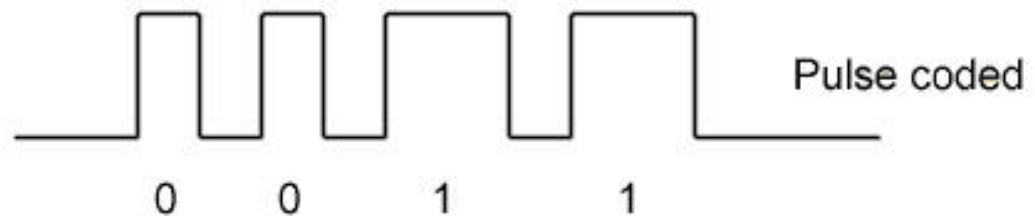
### **6.3.2. Design and Construction of long range infrared location emitter/receiver tags**

To improve on the design of the existing emitter/receiver infrared tags, I designed a daughter board, which, attached to the infrared tag, turns it into a long range infrared

emitter. Typically infrared I/O tags are built on a general-purpose prototyping board with a PIC microcontroller, serial I/O, and infrared input and serial output links. An example of hardware construction of infrared emitter/receiver tags are the iRX tags

[<http://www.media.mit.edu/~ayb/irx/irx2/>]. The PIC microcontroller is the heart of the board. It's a programmable microcontroller with 1K words of program memory and 68 bytes of general purpose RAM. It has 13 general purpose I/O ports. On the iRX board, five of these ports are dedicated (Serial In, Serial Out, Infrared Receive, Infrared LED, RED LED).

InfraRed (IR) data communication uses light waves in the Infra Red spectrum. The physical communication layer is formed by the emitter (a photo diode) that emits the signal and a receiver (another photo diode) that receives the signal. The light waves are modulated by the emitter at a frequency of 40 KHz. This is done in order to cut out other sources of IR such as electric lamps, etc. The data link layer is implemented by using binary pulses. The signal is pulse coded which means that the length of the pulse is varied to represent data.



**Figure 131. Example of pulse coded signal: PCW: the width of the pulse represents data.**

The idea I used for the construction of long range infrared emitter tags is that it is fine to overdrive an infrared diode beyond its specs, as long it is driven with very short pulses and it is later given the chance to rest a little (pause the transmission), after sending a signal. Overdriving here means having a current of 1A up to 1.5 A going through the infrared emitter diode even if the specs of the device rate it for a much less forward current, typically about 200mA. When driving the infrared diode with such a large current, the radiant emitted power increases linearly with the current that goes through the diode, and the range is greatly improved.

To do what I described above it suffices to use a power resistor in series with the infrared LED. However with such large currents involved it is impossible to drive the diode directly from one of the output pins of the microcontroller. I therefore use a power MOSFET transistor to drive the diode and I send the desired signal to the gate for the MOSFET from the microcontroller. To generate higher currents it is useful to have as large as possible voltage drop across the diode plus power resistor in series. I therefore decided to use the 9V iRX voltage supply as the positive voltage source for these components and grounded the other end of the circuit segment. I chose a power resistor value of 5 ohms to obtain a pulse forward current across the IR emitting diode of  $(9V - 2V \text{ of voltage drop across the diode}) / 5 \text{ ohms} = 1.4A$ . For efficiency, and to be able to actually drive the gate of the mosfet transistor between 9V and ground, I used a mosfet driver: the MAX 4420. Given a voltage supply of 9V and a square wave signal between 0

and 5V at the input, the MAX4420 reproduces the same input signal at the output, except that the output signal oscillates between 0 and 9V. The MAX4420 has also the added convenience of having two outputs and it therefore can drive two infrared emitting diodes at the same time for added range just with one single input signal from the microcontroller.

In the receiver tag, I used an infrared receiver module, the Panasonic PNA4613M00XD, which demodulates the incoming signal from the 40KHz carrier, and sends it to the output. The use of such a detector, rather than a regular receiver diode is highly recommended. Besides demodulation, it offers signal amplification and noise removal, it has a visible light cutoff resin over the detection pin diode to block visible light, and therefore gives the maximum reception distance. In addition, it requires little or no external components for operation. Of course it needs to be selected at the same spectral sensitivity and at the same wavelength of the selected infrared LED, which for the emitter tags is 940nm, and to operate at the exact frequency of the modulated data carrier, which for the SONY infrared protocol I use is 40KHz. The infrared receiver module is very sensitive to the value of the powering line: if it is not 5V exactly, the device malfunctions and causes noisy and shorter range detection. To avoid this problem I replaced the 78L05 voltage regulator of the iRX board with a better quality 5V regulator: the LT1121CZ-5, with a tantalum 1uF capacitor at the output. This low dropout regulator was chosen because it can supply 150ma of current, has reverse battery protection, and a low 0.4 volt dropout voltage. For the circuit design of the receiver I use a common NPN transistor, the 2N3904, to invert the signal and clamp it between 0 and 5V, which makes decoding a much easier task for the microcontroller.

The power consumption for the emitter daughter board is calculated by adding the power dissipated both by the LED and the power resistor which have respectively a voltage drop of 1.5V and 7.5V. The current in that circuit segment is 1.4A, as calculated above. Therefore for a driving square wave at 20% of its duty cycle, the emitter dissipates:  $9V \times 1.4A \times 0.5 \times 0.2 = 1.26W$ .

I tested range and directionality of the location identification sensor with a variety of high power infrared diodes. Assuming that the emitter tags generate a cone of infrared light, as shown in figures 135,136, I measured the height and diameter of the base of the cone both with one and two emitting diodes. I measured distance along a straight line, to test the maximum range. I measured the diameter of the emitted cone of infrared light at 6 feet [table 46], which corresponds to having the emitters placed in the museum with the other lights at about eleven feet from the floor, and assuming that people's head – where the infrared sensor is located – is at a (conservative) average height of 5 feet. The reader can verify from table 46 that a pair of OED-EL-1L2 diodes by Lumex reach up to 135 feet. The LN51L diode has the narrowest experimental angle of coverage of approximately 26° and the Lumex OED-EL-1L2 has the widest angle of coverage of 34°. The reason why the measured angle of coverage is much different from the one given by the manufacturer's specifications is that I overdrive the diode to emit a much higher radiance power than the value that was used to measure the angle of coverage by the manufacturer.

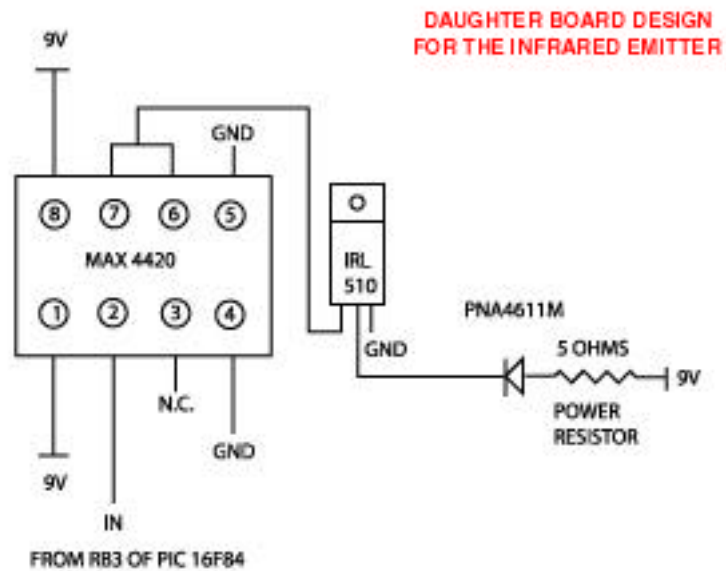


Figure 132. Circuit Diagram of the infrared emitter daughter board

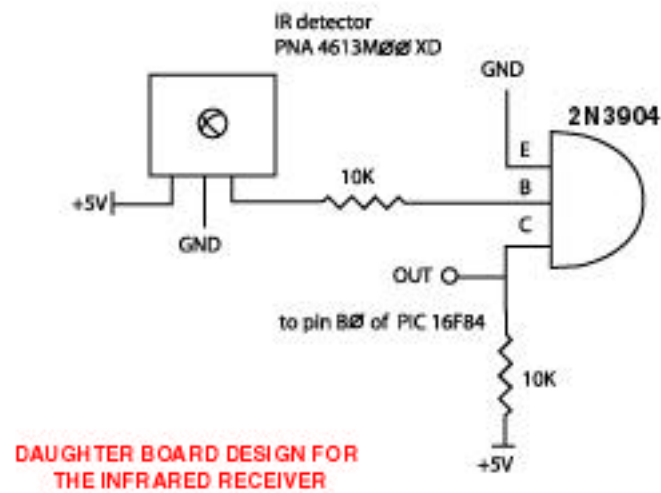
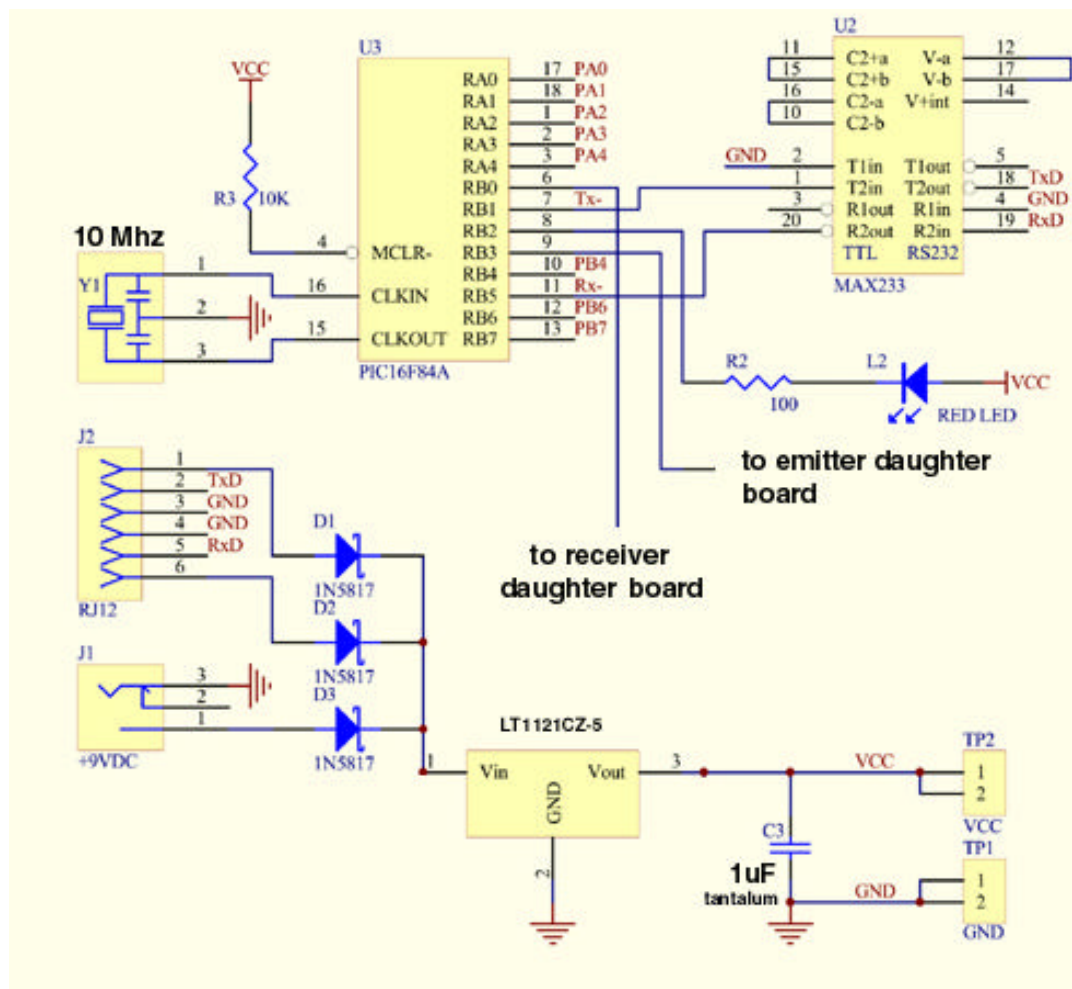
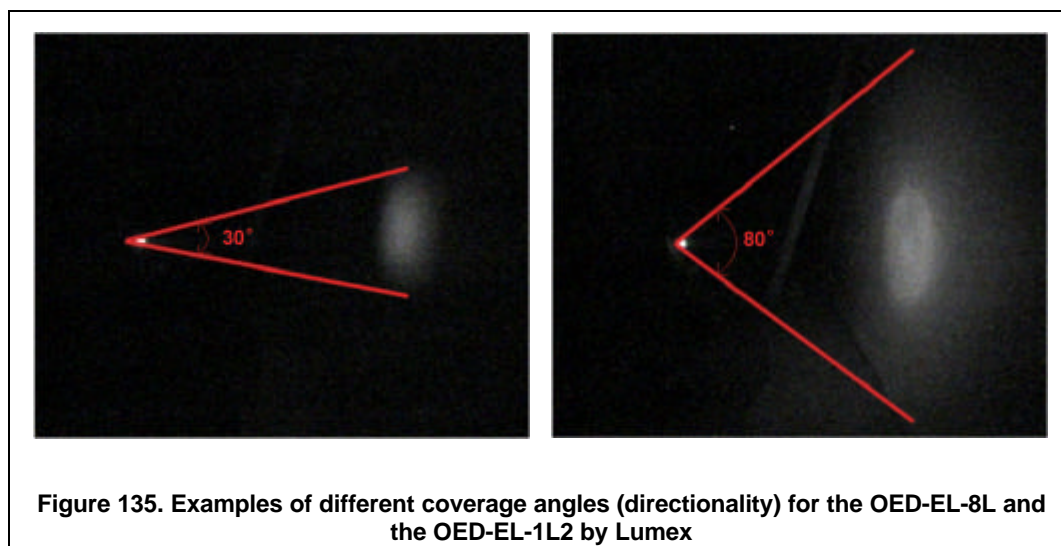


Figure 133. Circuit Diagram of the infrared receiver daughter board



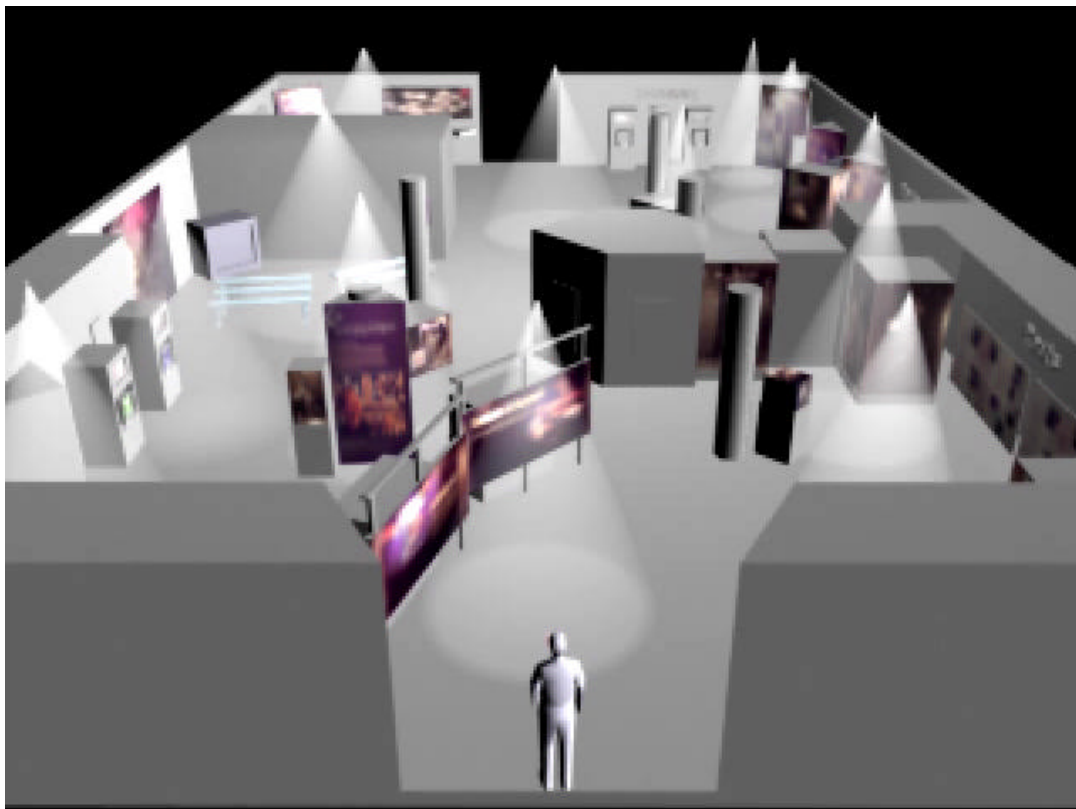
**Figure 134. Circuit Diagram of the infrared emitter/receiver mother board**





Diode Model	Pwr (mW)	Specs (theta)	$\Phi$ at 6 ft 1 diode (in)	$\Phi$ at 6 ft 2 diodes (in)	Theta exp	Distance 1 diode	Distance 2 diodes
LN 51L panasonic	6 mW	8	125	105	26	70 ft	100 ft
LN 166 panasonic	10 mW	20	173	187	32	95 ft	125 ft
OED-EL-1L2 LUMEX	25 mW	60	184	217	34	100 ft	135 ft
LN 66A panasonic	9 mW	25	139	182	30	90 ft	120 ft
OED-EL-8L LUMEX	3 mW	30	177	212	32	100 ft	130 ft

Table 46. Comparison of various commercially available infrared diodes. All infrared diodes operate at 940nm.



Figures 136. 3D model of MIT's *Robots and Beyond* exhibit which shows the cones of infrared light from the infrared location sensors

For transmission I used the standard SONY remote control infrared protocol which I modified slightly to transmit an ascii character made of eight bits, rather than the standard twelve bit: 5 for the address and 5 for the command, typically transmitted by SONY remote control units. According to the SONY infrared communication protocol, the light waves are modulated by the emitter at a frequency of 40 KHz. This is done in order to cut out noise from other sources of IR such as electric lamps, etc. The data is sent using pulse coding. I modified the standard protocol to send packets of 8 bits preceded by a header. Each packet varies in length between 100 and 160 milliseconds and is followed by 200ms of silence to allow the infrared diode to rest after being overdriven at 1.4A. The overall frequency of transmission of a location ID is therefore approximately 3Hz. The 40KHz carrier is generated only at 20% of its duty cycle, once again to preserve the lifetime of the overdriven infrared diode, as shown in figure 137.

The other specifications for this protocol are:

- Basic time period  $T = 550$  micro secs
- Header length =  $4T$  followed by  $T$  space
- 0 = Pulse with length  $T$  followed by space of length  $T$
- 1 = Pulse with length  $2T$  followed by space of length  $T$

The emitter location identification tags have been embedded inside standard light fixtures to allow the exhibit designer to easily place them in the museum, next to the regular museum lights, and using the same power rack as the regular museum spotlights [figures 140-145].

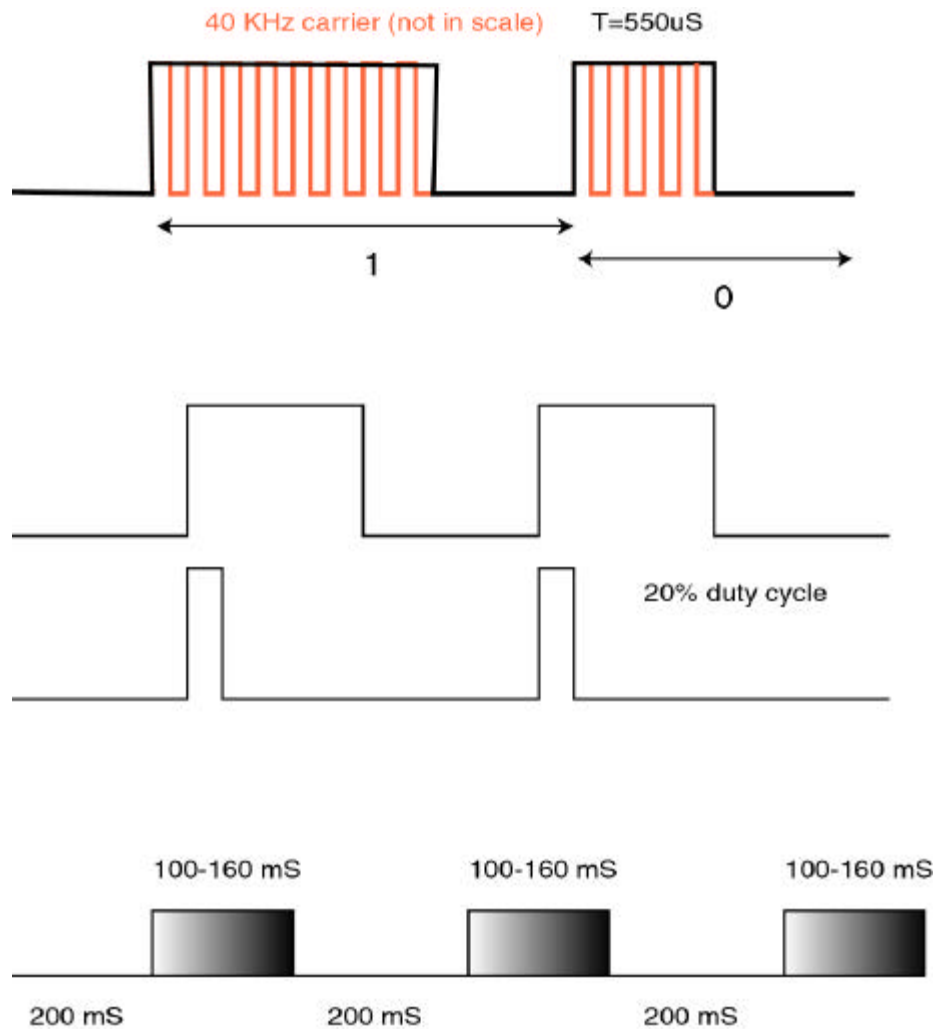


Figure 137. Pulse width modulation of the infrared location signal. The carrier is 40 KHz, generated at a 20 % duty cycle to preserve the life time of the infrared emitter diode. Below: it takes about 100-160 ms for the emitter to send an 8 bytes (characters) location signal, and after sending it, the software pauses for 200 ms, again to preserve the life time of the overdriven diode. The final location emission rate is therefore approximately 3 Hz.



Figure 138. Maximum range of the location sensor

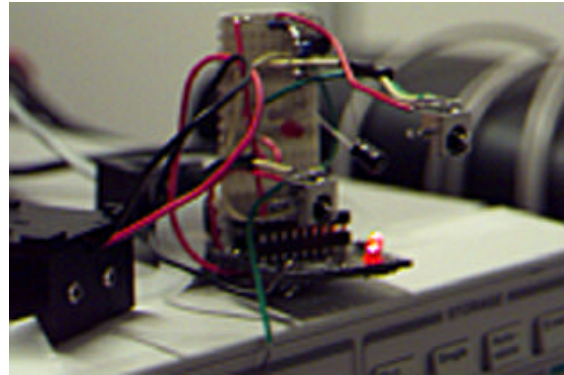


Figure 139. Location sensor: receiver



Figures 140, 141, 142, 143, 144, 145. Location sensor: emitter tags embedded inside light fixtures

## 6.4. The software

The museum wearable plays out an interactive documentary on the displayed artwork as video on the head mounted display augmented display glasses. The video is edited in small segments which vary in size from twenty seconds to one and half minute. A video server, written in C++ and DirectX 8, plays these clips and receives TCP/IP messages, representing which clip to play, from another program which weights the visitor's initial preferences which the information measured by the location ID sensors. This server-client architecture allows the programmer to easily add other client programs to the application, which communicate to the server information from other possible sources, such as sensors or cameras placed along the museum aisles, which measure the crowdedness of the galleries or how often a certain object has been visited. The client program reads IR data from the serial port and the server program does inference, content selection, and content playout using DirectX for full screen playout of the MPEG compressed clips. Using MPEG compression for the video clips allows for great space-saving on disk, as well as smooth full screen playout at video resolution.

This program uses HUGIN's software library [www.hugin.com] to perform real time probabilistic inference (see Chapter 4 for more details on the probability update algorithm used) on the basis of the information obtained from the infrared location tags and selects movie clips as a function of the visitor's estimated type.

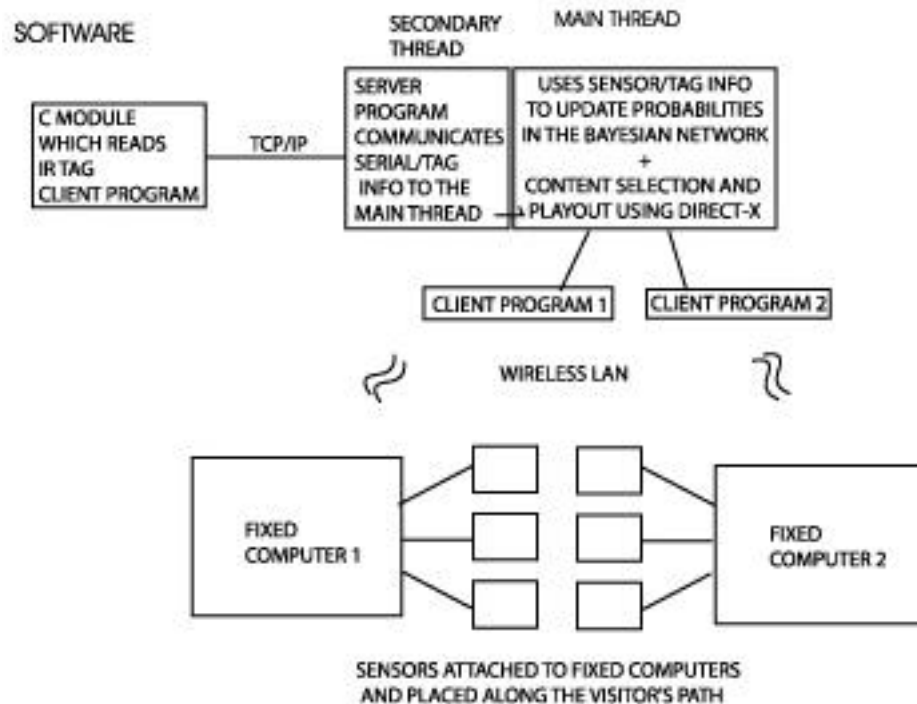


Figure 146. Diagram of the software architecture for the museum wearable.

## 6.5. The head mounted display





The size and weight of both the wearable's CPU and glasses are critical for a museum application. The augmented reality glasses cannot have a heavy and power hungry powering unit which requires frequent battery changes. Glasses also need to easily fit various people's head sizes, with annexed hair style, which is not an easy task. The wearable would be handed out to between ten and one hundred people a day, and therefore needs to be of robust assembly and easy to wear. Given that it is difficult to design one single head mounted display that fits all users, I have designed and assembled three different displays, which different styles and mounts. All these fashioned assemblies use commercial heads up displays, which I have thoroughly researched according to the following requirements. determining the most resourceful visual display, five characteristics are key to the model ideal in our application: *resolution*, *power consumption*, *purchase availability* and foremost, *ergonomics* with the inclusion or *adaptability with sensors*, *weight*, and *size*. I have considered four models for candidates: Sony's PML-S700 PC Glasstron, MicroOptical's CO-3, Daeyang's DH-4500MA, and Olympus M2 sold in the US at Tekgear retailers.

Resolution is vital for the museum visitor to be capable of reading text clearly especially while walking. VGA resolution provides the minimum resolution acceptable for text reading and enjoyable crisp images which do not produce eye-strain for viewing. An improvement, also in the visual field, is the superiority of having a partial see-through display rather than complete blockage of view by the display. In Sony's Glasstron, it is possible to remove one LCD producing monocular see-through view, while Daeyang's model is already monocular, yet not see-through.

RESOLUTION	COLUMNS	ROWS
SVGA	800	640
NTSC	720	480
VGA	640	480
QVGA	320	240





Table 47. Video resolution acronyms

The consumption of power limits the operation of the wearable unit. Addition of devices and sensors decrease the time of functionality before another battery is required. Therefore a visual display with the minimum power consumption would be optimal. For a theoretical mobile operation of an hour and a half, with the inclusion of the computer and sensors, the desire model should consume less than 6 Watts. Sony's Glasstron with exceptional resolution requires a power input above the battery limit.

Model		Resolution (H)*(V)	Number of Pixels	Opaque or Translucent
Sony Glasstron PLM-S700		832*624	1.55 million	Translucent w/o LCD
Daeyang DH-4500MA		SVGA	1.44 million	1 eye opaque
Olympus Tekgear M2		SVGA	480,000	Translucent
MicroOptical		VGA	-----	

**Table 48. Comparison of commercially available head mounted displays, for their video resolution**


The DH-4500MA, was announced to be available beginning of June 2001, with its superior resolution at a low cost, it was a probable winner for the museum wearable project. Daeyang recently has announced it will delay its introduction until the near 3rd quarter of the year, too late for the completion of this project, leaving the Olympus M2, and the MicroOptical.CO-3 as the final candidates.

Model		Power Consumption	Voltage DC
Sony Glasstron PLM-S700		10 W	8.4
Daeyang DH-4500MA		4 W	5
Olympus Tekgear M2		2 W	5
MicroOptical		3 W	7.2

**Table 49. Comparison of commercially available head mounted displays, for their power consumption**



Ergonomics plays a key role for the final elimination. In pursuing models with higher resolutions, some adjustments must be made for our application. The ideal model will have only one LCD, leaving one eye to observe the museum media. Sensor adjustments should be as easy as possible, the display needs to be lightweight, adaptable to different head sizes, and comfortable to the user. The following figure shows some adjustments on the final candidates to meet our requirements.

Model		Weight	Required adjustments
Olympus Tekgear M2		210g	Addition of sensors
MicroOptical		40g(w/o box)	Addition of headphones and sensors

**Table 50. Required adjustments for use with the museum wearable for the selected displays.**

*Cost* became the final criteria to select the head mounted display. The color VGA resolution MicroOptical costs half the price of the Olympus M2 (\$2,500 vs \$5,000) and it was therefore selected to be the display of choice for the museum wearable.

Below is a description of the product design study I carried out on three different fashioning of the head mounted display so that it can adapt to the different needs and head sizes of museum visitors.

### **The common fashion display**

The common fashion display features an augmented reality display, which joins together a lightweight VGA resolution color display from the MicroOptical corporation, and a commercial high quality sturdy set of headphones. The two are joined rigidly by a metallic mount, which is attached to the side of the headphones and has a slot designed to accommodate the clip of the MicroOptical display. I designed this mount after several optimization iterations, so that it does not have pointy parts that could hurt the wearer and it is flexible so that it can bend slightly towards to face of the wearer to align the display along the visual axis of the visitor. It has been created in stainless steel with the user of a water jet cutter machine.

The choice of headphones is a compromise between quality and weight. To avoid having a display that bounces as the visitor moves along the exhibit the headphones need to be somewhat sturdy, although not as heavy as professional headphones, which would be uncomfortable for the common visitor to wear. The infrared sensor is placed in the center of the headband, and it is connected by a thin wire to the circuitry that reads the data and sends it to the serial port of the computer. This circuitry, described in the

following paragraph, is located inside the small shoulder pack which also holds the computer.



Figures 147, 148, 149, 150, 151. Product design and assembly of the common fashion display

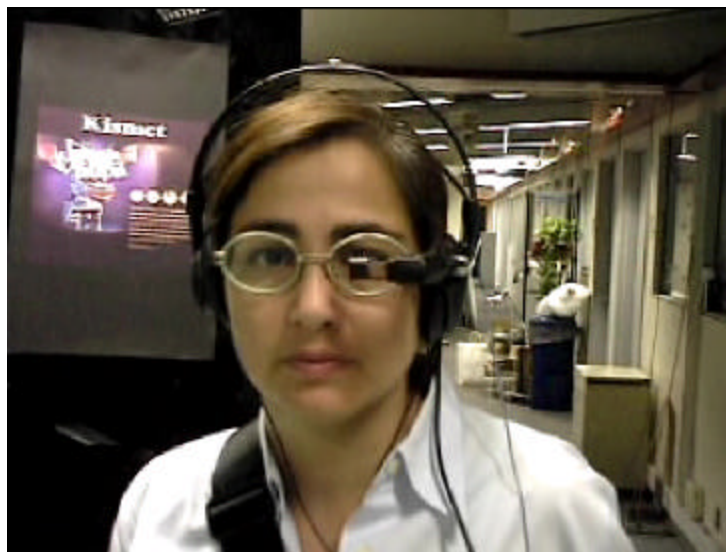
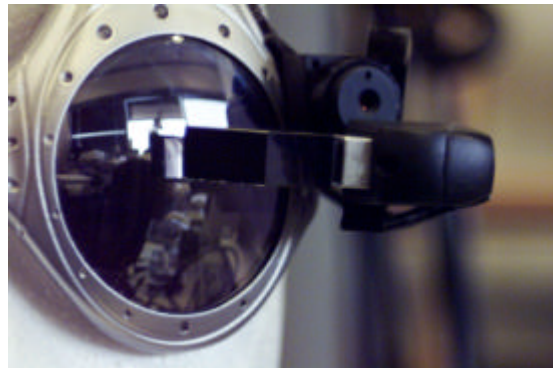
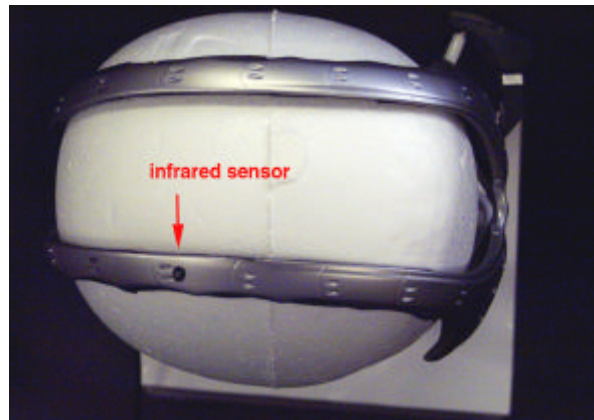


Figure 152. Museum visitor wearing the common fashion display

### **The high fashion display**

This design is a provocative stylish mount, mainly intended for visitors with a strong sense of aesthetics, and suitable for use in wearable fashion shows, to promote a non-nerdy and high fashionable wear of augmented reality displays for the large public. The MicroOptical augmented reality display is here rigidly mounted to a pair of Oakley “over the top” glasses, as illustrated in the figures 153-156. This slick mount was especially designed for Olympic athletes who need to have glasses which do not bounce or move while they perform. The infrared sensor is also located in the center of the headband, as in the previous design.



**Figures 153, 154, 155, 156. High fashion display: MicroOptical wearable display mounted onto Oakley over-the-top glasses with study for camera placement**

### **The old fashion display**

I propose this design for visitors who feel uncomfortable wearing an augmented reality display all along the length of their visit at the museums. These are potentially visitors with special needs, elderly people, or more cautious visitors who would like to try a less immersive augmented reality device, than the two previously illustrated. The old fashion display imitates the design of old opera glasses or binoculars, that people would wear momentarily to see details or close-ups of a theatrical or musical performance. It is made by a large pair of glasses, from which we have removed one lateral arm. The remaining arm is rigidly joined with a stainless steel joint which provides a support to attach the MicroOptical augmented reality display, as well as a one-ear headphone. The infrared sensor is located in between the lenses of the glasses.



**Figures 157, 158. The old fashion display**

## Chapter 7

# Results and Evaluation

### 7.1. Model Validation

Classifying people's behavior using the dynamic Bayesian network described in Chapter 4 can be done on the sole basis of the expert's opinion, which assigns numerical values to the conditional probability tables of the network. This results in an expert system, and in this case the network reflects the subjective opinion and domain knowledge of the people who designed it. In some cases performing probability update on such models leads to satisfying results, such as for systems that model the specialist's medical knowledge in decision support systems which help determine whether further tests or investigation are needed to assess the health state of a patient.

When a database of cases is available, we have the opportunity to perform out-of-sample testing to validate that the model has any predictive power, given the data. This is important in the case of a museum exhibit in which the public's behavior cannot necessarily be correctly modeled by the curator's expertise, because of the changing nature of the shown artwork, and the public's needs.

According to the problem that is defined there are various techniques for data analysis which combine prior knowledge with data to produce improved knowledge. In some cases we may need to learn the topology of the network, in others, the topology is given and we need to learn the parameters of the network. The four possible cases and techniques, taken from Murphy and Mian [Murphy and Mian, 1999], are described in the table 51. Heckerman [1999] provides an introduction to some of the issues involved in learning with a Bayesian network.

Structure	Observability	Method
Known	Full	Sample Statistics
Known	Partial	EM or gradient ascent
Unknown	Full	Search through model space
Unknown	Partial	Structural EM

**Table 51. Learning methods depending on what is already known about the problem. Taken from Murphy and Mian [1999].**

### 7.1.1. Learning from the data

Specifically for this research, I made the assumption that the structure of the network is correct and I used the tracking data on visitor's path and length of stops at the museum to learn the conditional probability tables (parameters) for the nodes of the network. I grouped the visitor tracking data gathered at the museum in two groups: a group from which I trained the proposed Bayesian network (i.e. I learned the parameters of the network from the data) and a control group on which I performed out-of-sample testing to validate the predictive power of the model. For robustness I split the data in a training group and a control group in three different ways, repeated the learning procedure three times, and compared the obtained results. The training and test groups are listed in tables 55-61.

The Bayesian network on which I perform learning [figure 62] has three nodes per time slice: the *object* and *location* nodes are dynamic, i.e. they are repeated at each time slice, whereas the *visitor* node is a static node, as I assume that the visitor type does not change during the visit at the museum.

The input to the network are the location nodes, which express how much time the visitors spend at each object. Rather than having location nodes with continuous values, which would require more sophisticated probability updates techniques, it is more convenient to have discrete location nodes with a few states, such as "short", "long" stop duration, or "skip" object. Section 7.1.2. describes how, mathematically, stop durations are labeled as skip/short/long. This section focuses on model validation and out-of-sample testing, and assumes that the input data has already been correctly classified.

The nodes of the network have discrete states as below [tables 52,53,54]:

Location (L)	Visitor (V)	Object (O)	Object' (O')
Skip	Busy	Visited	Neutral
Short	Greedy	Not Visited	Interesting
Long	Selective		Uninteresting

For simplicity, in this first phase of this research, I have assigned only two trivial states to the object node: Visited and Not Visited. A more accurate description could include three states for the Object nodes: Neutral, Interesting, Uninteresting, based on the number of visits they receive and length of stops. Given that the system is learning the conditional probability table of the location nodes, given the visitor and object nodes, i.e.  $p(L|O,V)$ , in the  $p(L|O',V)$  case it would have to learn 27 parameters (3 location states x 3 object states x 3 visitor states) instead of 18 (3x2x3).

<i>visitor priors</i>			<i>object 1 priors</i>			<i>object 2-12 CPT</i>		
<b>Busy</b>	0.333		<b>visited</b>	0.5		<b>visited</b>	0.5	0.5
<b>Greedy</b>	0.333		<b>not visited</b>	0.5		<b>not visited</b>	0.5	0.5
<b>Selective</b>	0.333							



<i>location nodes Conditional Probability Table</i>						
<i>type</i>	<b>busy</b>		<b>Greedy</b>		<b>selective</b>	
<i>object 1</i>	visited	not visited	visited	not visited	visited	not visited
<b>skip</b>	0.2	1	0.1	1	0.4	1
<b>short</b>	0.7	0	0.1	0	0.2	0
<b>long</b>	1	0	0.8	0	0.4	0

Tables 52,53,54. Initial (before learning) probability tables of the nodes of the dynamic Bayesian network.

To learn the 9  $p(L|O,V)$  parameters (only 9 to learn for the visited nodes) I applied the Expectation Maximization algorithm. The Expectation-Maximization (EM) algorithm is a broadly applicable approach to the iterative computation of maximum likelihood (ML) estimates, useful in a variety of incomplete-data problems. On each iteration of the EM algorithm, there are two steps, called the expectation step or the E-step and the maximization step or the M-step. Because of this, the algorithm is called the EM algorithm. This name was given by Dempster, Laird, and Rubin [Dempster, Laird, and Rubin, 1977] in their fundamental paper. On each iteration of the EM algorithm, there are two steps, called the expectation step or the E-step and the maximization step or the M-step. Because of this, the algorithm is called the EM algorithm.

The basic idea of the EM algorithm is to associate with the given incomplete data problem, a complete-data problem for which ML estimation is computationally more tractable; for instance, the complete-data problem chosen may yield a closed-form solution to the maximum likelihood estimate (MLE) or may be amenable to MLE computation with a standard computer package. The methodology of the EM algorithm then consists in reformulating the problem in terms of this more easily solved complete data problem, establishing a relationship between the likelihoods of these two problems, and exploiting the simpler MLE computation of the complete data problem in the M-step of the iterative computing algorithm.

More specifically, let's assume we wish to estimate some set of parameters  $\theta$  that describe an underlying probability distribution, given only the observed portion of the full data produced by this distribution. Let  $X = \{x_1, x_2, \dots, x_m\}$  denote the observed data in a set of  $m$  independently drawn instances, let  $Z = \{z_1, z_2, \dots, z_m\}$  denote the unobserved data

let  $Y = X \cup Z$  denote the full data. We use  $h$  to indicate the current hypothesized values of the parameters  $\theta$ , and  $h'$  to denote the revised hypothesis that is estimated on each iteration of the EM algorithm. The EM algorithm searches for the maximum likelihood hypothesis  $h'$  by seeking the  $h'$  that maximizes:  $E[\ln p(Y|h')]$ . It uses its current hypothesis  $h$  in place of the actual parameters  $\theta$  to estimate the distribution governing  $Y$ . Let us define a function  $Q(h|h)$  that gives  $E[\ln p(Y|h)]$  as a function of  $h'$ , under the assumption that  $\theta = h$  and given the observed portion of the full data  $Y$ :  $Q(h|h) = E[\ln p(Y|h') | h, X]$ . The EM can be described as:

Step 1. *Estimation (E) step*: Calculate  $Q(h|h)$  using the current hypothesis  $h$  and the observed data  $X$  to estimate the probability distribution over  $Y$ .

$$Q(h|h) \leftarrow E[\ln p(Y|h') | h, X]$$



Step 2. *Maximization (M) step*: Replace hypothesis  $h$  by the hypothesis  $h'$  that maximizes the  $Q$  function.

$$h \leftarrow \underset{h'}{\operatorname{argmax}} Q(h|h)$$

The results of EM parameter learning for the three training groups in tables 55, 57, 59, are given in table 61.

	Intro-1	Lisp-2	MinskyAr m-3	RoboAr m-4	Falcon-5	Phantom- 6	CogsHea d-7	Quad-8	uniroo-9	Dext Arm-10	Kismet- 11	Baby Doll-12	TYPE
1	skip	skip	long	short	long	long	short	long	long	long	long	short	greedy
2	short	long	long	long	long	long	long	long	long	skip	short	long	greedy
3	short	skip	short	short	short	short	skip	skip	skip	skip	long	skip	busy
4	short	short	short	short	short	short	long	short	skip	short	skip	long	busy
5	short	short	short	short	short	short	short	long	long	short	short	short	busy
6	short	short	long	skip	skip	skip	short	short	short	short	long	short	busy
7	short	skip	skip	skip	short	skip	short	short	short	long	short	short	busy
8	short	short	short	short	short	short	skip	short	short	short	long	skip	busy
9	short	short	short	skip	skip	skip	short	long	short	short	skip	skip	busy
10	short	short	short	short	skip	short	skip	skip	skip	short	short	short	busy
11	short	long	long	long	short	skip	long	skip	skip	skip	skip	short	sctv
12	short	skip	short	short	short	skip	short	skip	skip	long	long	skip	sctv
13	skip	long	short	long	long	short	long	skip	skip	short	skip	long	sctv
14	skip	skip	skip	skip	long	short	short	skip	short	long	long	short	sctv
15	long	long	long	long	long	short	short	skip	skip	skip	long	skip	sctv
16	skip	short	skip	skip	long	long	skip	skip	skip	long	long	long	sctv
17	long	long	long	long	short	long	short	short	short	skip	short	skip	sctv
18	long	short	skip	short	skip	skip	short	long	long	short	skip	skip	sctv
19	short	long	short	long	skip	skip	skip	long	short	long	skip	long	sctv
20	short	long	short	short	long	long	skip	skip	skip	skip	skip	short	sctv
21	short	long	short	short	skip	skip	short	skip	skip	short	skip	long	sctv
22	short	short	short	long	short	long	long	skip	skip	skip	long	short	sctv
23	short	long	short	skip	short	short	short	skip	long	skip	long	skip	sctv
24	short	long	short	long	long	long	short	skip	skip	skip	skip	skip	sctv
25	short	long	long	long	long	skip	skip	skip	short	long	skip	skip	sctv

Table 55. Training group 1

	Intro-1	Lisp-2	MinskyAr m-3	RoboAr m-4	Falcon-5	Phantom- 6	CogsHea d-7	Quad-8	uniroo-9	Dext Arm-10	Kismet- 11	Baby Doll-12	TYPE
1	skip	short	long	short	long	long	skip	long	long	long	long	short	greedy
2	short	short	short	short	short	short	short	short	short	skip	long	skip	busy
3	short	skip	skip	skip	short	skip	skip	short	short	short	long	short	busy
4	short	short	short	skip	skip	short	short	skip	skip	skip	long	short	busy
5	short	skip	skip	skip	short	short	short	short	short	short	skip	short	busy
6	short	long	short	short	long	long	long	short	short	short	short	short	busy
7	short	long	short	short	long	short	long	short	short	short	long	skip	busy
8	skip	short	short	short	short	skip	short	skip	short	short	short	short	busy
9	short	short	short	short	long	short	short	long	long	short	skip	skip	busy

10	short	skip	long	long	long	long	skip	skip	long	short	long	skip	slctv
11	short	skip	long	short	long	short	skip	skip	skip	short	long	short	slctv
12	skip	skip	short	long	long	long	skip	long	short	long	skip	short	slctv
13	long	long	long	skip	skip	short	short	long	short	short	skip	skip	slctv
14	short	short	short	skip	skip	short	long	short	long	skip	long	skip	slctv
15	long	skip	skip	skip	skip	skip	short	long	long	long	short	long	slctv
16	skip	skip	short	long	skip	skip	long	short	short	short	skip	long	slctv
17	short	short	long	long	long	long	skip	skip	long	skip	long	skip	slctv
18	short	skip	short	short	skip	long	long	long	skip	long	short	skip	slctv
19	short	skip	short	skip	long	short	long	short	long	skip	long	skip	slctv
20	short	long	long	long	long	skip	skip	skip	short	short	long	skip	slctv
21	short	skip	long	skip	short	skip	short	skip	short	long	long	short	slctv
22	skip	long	short	skip	long	skip	long	long	long	long	skip	skip	slctv
23	short	long	short	long	long	short	short	long	skip	skip	skip	skip	slctv
24	short	short	skip	short	long	short	long	skip	long	skip	long	skip	slctv
25	short	skip	skip	long	short	short	long	long	short	short	skip	skip	slctv

Table 56. Test group 1

	Intro-1	Lisp-2	MinskyAr m-3	RoboAr m-4	Falcon-5	Phantom- 6	CogsHea d-7	Quad-8	uniroo-9	Dext Arm-10	Kismet- 11	Baby Doll-12	TYPE
1	short	long	long	long	long	long	long	long	long	skip	short	long	greedy
2	short	short	long	skip	skip	skip	short	short	short	short	long	short	busy
3	short	skip	skip	skip	short	skip	short	short	short	long	short	short	busy
4	short	skip	skip	skip	short	skip	skip	short	short	short	long	short	busy
5	short	short	short	short	short	short	skip	short	short	short	long	skip	busy
6	short	short	short	skip	skip	skip	short	long	short	short	skip	skip	busy
7	short	long	short	short	long	short	long	short	short	short	long	skip	busy
8	short	short	short	short	skip	short	skip	skip	skip	short	short	short	busy
9	short	short	short	skip	skip	short	short	skip	skip	skip	long	short	busy
10	long	short	skip	short	skip	skip	short	long	long	short	skip	skip	slctv
11	short	short	long	long	long	long	skip	skip	long	skip	long	skip	slctv
12	short	long	short	long	skip	skip	skip	long	short	long	skip	long	slctv
13	short	long	short	long	long	short	short	long	skip	skip	skip	skip	slctv
14	short	long	short	short	long	long	skip	skip	skip	skip	skip	short	slctv
15	short	skip	short	short	skip	long	long	long	skip	long	short	skip	slctv
16	short	short	skip	short	long	short	long	skip	long	skip	long	skip	slctv
17	short	long	short	short	skip	skip	short	skip	skip	short	skip	long	slctv
18	skip	long	short	skip	long	skip	long	long	long	long	skip	skip	slctv
19	short	short	short	long	short	long	long	skip	skip	skip	long	short	slctv
20	short	skip	short	skip	long	short	long	short	long	skip	long	skip	slctv
21	short	long	short	skip	short	short	short	skip	long	skip	long	skip	slctv
22	short	skip	long	skip	short	skip	short	skip	short	long	long	short	slctv
23	short	long	short	long	long	long	short	skip	skip	skip	skip	skip	slctv
24	short	long	long	long	long	skip	skip	skip	short	short	long	skip	slctv
25	short	long	long	long	long	skip	skip	skip	short	long	skip	skip	slctv

Table 57. Training group 2

	Intro-1	Lisp-2	MinskyAr m-3	RoboAr m-4	Falcon-5	Phantom- 6	CogsHea d-7	Quad-8	uniroo-9	Dext Arm-10	Kismet- 11	Baby Doll-12	TYPE
1	skip	skip	long	short	long	long	short	long	long	long	long	short	greedy
2	skip	short	long	short	long	long	skip	long	long	long	long	short	greedy
3	skip	short	short	short	short	skip	short	skip	short	short	short	short	busy
4	short	skip	short	short	short	short	skip	skip	skip	skip	long	skip	busy
5	short	short	short	short	short	short	short	short	short	skip	long	skip	busy
6	short	short	short	short	short	short	long	short	skip	short	skip	long	busy
7	short	short	short	short	long	short	short	long	long	short	skip	skip	busy
8	short	skip	skip	skip	short	short	short	short	short	short	skip	short	busy
9	short	short	short	short	short	short	short	long	long	short	short	short	busy
10	short	long	short	short	long	long	long	short	short	short	short	short	busy
11	short	skip	long	long	long	long	skip	skip	long	short	long	skip	sctv
12	short	long	long	long	short	skip	long	skip	skip	skip	skip	short	sctv
13	skip	skip	short	long	skip	skip	long	short	short	short	skip	long	sctv
14	short	skip	short	short	short	skip	short	skip	skip	long	long	skip	sctv
15	short	skip	long	short	long	short	skip	skip	skip	short	long	short	sctv
16	skip	long	short	long	long	short	long	skip	skip	short	skip	long	sctv
17	long	long	long	skip	skip	short	short	long	short	short	skip	skip	sctv
18	skip	skip	skip	skip	long	short	short	skip	short	long	long	short	sctv
19	long	long	long	long	long	short	short	skip	skip	skip	long	skip	sctv
20	skip	skip	short	long	long	long	skip	long	short	long	skip	short	sctv
21	skip	short	skip	skip	long	long	skip	skip	skip	long	long	long	sctv
22	long	skip	skip	skip	skip	skip	short	long	long	long	short	long	sctv
23	long	long	long	long	short	long	short	short	short	skip	short	skip	sctv
24	short	skip	skip	long	short	short	long	long	short	short	skip	skip	sctv
25	short	short	short	skip	skip	short	long	short	long	skip	long	skip	sctv

Table 58. Test group 2

	Intro-1	Lisp-2	MinskyAr m-3	RoboAr m-4	Falcon-5	Phantom- 6	CogsHea d-7	Quad-8	uniroo-9	Dext Arm-10	Kismet- 11	Baby Doll-12	TYPE
1	skip	skip	long	short	long	long	short	long	long	long	long	short	greedy
2	skip	short	short	short	short	skip	short	skip	short	short	short	short	busy
3	short	skip	short	short	short	short	skip	skip	skip	skip	long	skip	busy
4	short	short	short	short	short	short	short	short	short	skip	long	skip	busy
5	short	short	short	short	short	short	long	short	skip	short	skip	long	busy
6	short	short	short	short	long	short	short	long	long	short	skip	skip	busy
7	short	skip	skip	skip	short	short	short	short	short	short	skip	short	busy
8	short	short	short	short	short	short	short	long	long	short	short	short	busy
9	short	long	short	short	long	long	long	short	short	short	short	short	busy
10	short	skip	long	long	long	long	skip	skip	long	short	long	skip	sctv
11	short	long	long	long	short	skip	long	skip	skip	skip	skip	short	sctv
12	skip	skip	short	long	skip	skip	long	short	short	short	skip	long	sctv
13	short	skip	short	short	short	skip	short	skip	skip	long	long	skip	sctv
14	short	skip	long	short	long	short	skip	skip	skip	short	long	short	sctv
15	skip	long	short	long	long	short	long	skip	skip	short	skip	long	sctv
16	long	long	long	skip	skip	short	short	long	short	short	skip	skip	sctv
17	skip	skip	skip	skip	long	short	short	skip	short	long	long	short	sctv
18	long	long	long	long	long	short	short	skip	skip	skip	long	skip	sctv
19	skip	skip	short	long	long	long	skip	long	short	long	skip	short	sctv

20	skip	short	skip	skip	long	long	skip	skip	skip	long	long	long	slctv
21	long	skip	skip	skip	skip	skip	short	long	long	long	short	long	slctv
22	long	long	long	long	short	long	short	short	short	skip	short	skip	slctv
23	short	skip	skip	long	short	short	long	long	short	short	skip	skip	slctv
24	short	short	short	skip	skip	short	long	short	long	skip	long	skip	slctv
25	long	short	skip	short	skip	skip	short	long	long	short	skip	skip	slctv

Table 59. Training group 3

	Intro-1	Lisp-2	MinskyArm-3	RoboArm-4	Falcon-5	Phantom-6	CogsHead-7	Quad-8	uniroo-9	Dext Arm-10	Kismet-11	Baby Doll-12	
1	skip	short	long	short	long	long	skip	long	long	long	long	short	greedy
2	short	long	long	long	long	long	long	long	long	skip	short	long	greedy
3	short	short	long	skip	skip	skip	short	short	short	short	long	short	busy
4	short	skip	skip	skip	short	skip	short	short	short	long	short	short	busy
5	short	skip	skip	skip	short	skip	skip	short	short	short	long	short	busy
6	short	short	short	short	short	short	skip	short	short	short	long	skip	busy
7	short	short	short	skip	skip	skip	short	long	short	short	skip	skip	busy
8	short	long	short	short	long	short	long	short	short	short	long	skip	busy
9	short	short	short	short	skip	short	skip	skip	skip	short	short	short	busy
10	short	short	short	skip	skip	short	short	skip	skip	skip	long	short	busy
11	short	short	long	long	long	long	skip	skip	long	skip	long	skip	slctv
12	short	long	short	long	skip	skip	skip	long	short	long	skip	long	slctv
13	short	long	short	long	long	short	short	long	skip	skip	skip	skip	slctv
14	short	long	short	short	long	long	skip	skip	skip	skip	skip	short	slctv
15	short	skip	short	short	skip	long	long	long	skip	long	short	skip	slctv
16	short	short	skip	short	long	short	long	skip	long	skip	long	skip	slctv
17	short	long	short	short	skip	skip	short	skip	skip	short	skip	long	slctv
18	skip	long	short	skip	long	skip	long	long	long	long	skip	skip	slctv
19	short	short	short	long	short	long	long	skip	skip	skip	long	short	slctv
20	short	skip	short	skip	long	short	long	short	long	skip	long	skip	slctv
21	short	long	short	skip	short	short	short	skip	long	skip	long	skip	slctv
22	short	skip	long	skip	short	skip	short	skip	short	long	long	short	slctv
23	short	long	short	long	long	long	short	skip	skip	skip	skip	skip	slctv
24	short	long	long	long	long	skip	skip	skip	short	short	long	skip	slctv
25	short	long	long	long	long	skip	skip	skip	short	long	skip	skip	slctv

Table 60. Test group 3

NEW LEARNED conditional probability table, train group 1				Original conditional probability table			
	skip	short	long		skip	short	long
Busy	0.27	0.63	0.1	Busy	0.2	0.7	0.1
Greedy	0.13	0.21	0.66	Greedy	0.1	0.1	0.8
Selective	0.37	0.3	0.33	Selective	0.4	0.2	0.4

NEW LEARNED conditional probability table, train group 2				Original conditional probability table			
	skip	short	long		skip	short	long
Busy	0.3	0.59	0.11	Busy	0.2	0.7	0.1
Greedy	0.08	0.17	0.75	Breedy	0.1	0.1	0.8
Selective	0.36	0.3	0.34	Selective	0.4	0.2	0.4

NEW LEARNED conditional probability table, train group3				Original conditional probability table			
	skip	short	long		skip	short	long
Busy	0.2	0.67	0.13	Busy	0.2	0.7	0.1
Greedy	0.17	0.25	0.58	Greedy	0.1	0.1	0.8
Selective	0.36	0.3	0.34	Selective	0.4	0.2	0.4

Tables 61. New parameters learned from the visitor tracking data using the EM algorithm.

I used the new learned parameters to test how well the Bayesian network performs in identifying the other half of the tracked user types. In each case I took the original set of 50 visitor tracking data, and split it randomly in two parts, each composed of 25 subjects. I used the first half of the subjects as training data, and the remaining half as test data. For robustness I performed this operation three times, each time with a different random subdivision of the original data set in half. I then, for each of the three learned conditional probability tables for the location nodes, substituted the original conditional probability table with the new learned parameters. Then, for each of the 25 visitor data (row) in the test group, I introduced the stop duration for the 12 tracked objects as evidence in the network, and calculated the posterior probability on the visitor nodes. I compared the visitor's busy/greedy/selective state with the highest probability, with the label assigned to the visitor behavior in the test file. When the original labeled data coincided with the posterior with the highest probability I considered this a 'success', otherwise a 'miss'. For the three test cases, each made of 25 cases, I obtained respectively 25, 24, and 25 successes, which are listed in tables 64, 65, 66. There was only one miss, with test data as below, which the network classified as 'busy' while it was labeled 'selective'.

short	skip	short	short	short	skip	short	skip	skip	long	long	skip	selective
-------	------	-------	-------	-------	------	-------	------	------	------	------	------	-----------

The percentage of skips, short, and long stops for this test case is respectively:

%-busy	%-greedy	%-slctv
0.42	0.42	0.16

The absolute classification errors for the second test data group are:

E-busy	E-greedy	E-slctv
0.33	1.16	0.34

NEW LEARNED conditional probability table, train group 2			
	skip	short	long
Busy	0.3	0.58	0.12
Greedy	0.08	0.17	0.75
Selective	0.36	0.3	0.34

Table 62. New learned probability table, for train group 2

As one can observe from the error table, the misclassified visitor had therefore a behavior at the dividing boundary between busy and selective, as they have very close absolute errors. I would attribute the error to the human who observed the visitors' behavior at the museum, and labeled that particular's visitor behavior wrongly because of the ambiguity of that test case.

Given the high success rate of this learning/test procedure, which can be quantified as  $74/75=0.987$ , for further visitor classification with the given Bayesian network, I have performed EM learning on all 50 visitors obtaining the final learned parameters in table 63:

<i>Final Conditional Probability Table <math>p(L O,V)</math></i>			
	<b>skip</b>	<b>short</b>	<b>long</b>
<b>busy</b>	0.25	0.625	0.125
<b>greedy</b>	0.14	0.22	0.64
<b>selective</b>	0.36	0.3	0.34

Table 63. New learned probability table, for all 50 visitor data.

<i>test1: correct=25 incorrect=0</i>			
1	yes	result=greedy	original=greedy
2	yes	result=busy	original=busy
3	yes	result=busy	original=busy
4	yes	result=busy	original=busy
5	yes	result=busy	original=busy
6	yes	result=busy	original=busy
7	yes	result=busy	original=busy
8	yes	result=busy	original=busy
9	yes	result=busy	original=busy
10	yes	result=selective	original=selective
11	yes	result=selective	original=selective
12	yes	result=selective	original=selective
13	yes	result=selective	original=selective
14	yes	result=selective	original=selective
15	yes	result=selective	original=selective
16	yes	result=selective	original=selective
17	yes	result=selective	original=selective
18	yes	result=selective	original=selective
19	yes	result=selective	original=selective
20	yes	result=selective	original=selective
21	yes	result=selective	original=selective
22	yes	result=selective	original=selective
23	yes	result=selective	original=selective
24	yes	result=selective	original=selective
25	yes	result=selective	original=selective

Table 64. Test results for group 1.

<i>test2: correct=24 incorrect=1</i>			
1	yes	result=greedy	original=greedy
2	yes	result=greedy	original=greedy
3	yes	result=busy	original=busy
4	yes	result=busy	original=busy

5	yes	result=busy	original=busy
6	yes	result=busy	original=busy
7	yes	result=busy	original=busy
8	yes	result=busy	original=busy
9	yes	result=busy	original=busy
10	yes	result=busy	original=busy
11	yes	result=selective	original=selective
12	yes	result=selective	original=selective
13	yes	result=selective	original=selective
14	no	result=busy	original=selective
15	yes	result=selective	original=selective
16	yes	result=selective	original=selective
17	yes	result=selective	original=selective
18	yes	result=selective	original=selective
19	yes	result=selective	original=selective
20	yes	result=selective	original=selective
21	yes	result=selective	original=selective
22	yes	result=selective	original=selective
23	yes	result=selective	original=selective
24	yes	result=selective	original=selective
25	yes	result=selective	original=selective

Table 65. Test results for group 2.

<i>test3: correct=25 incorrect=0</i>			
1	yes	result=greedy	original=greedy
2	yes	result=greedy	original=greedy
3	yes	result=busy	original=busy
4	yes	result=busy	original=busy
5	yes	result=busy	original=busy
6	yes	result=busy	original=busy
7	yes	result=busy	original=busy
8	yes	result=busy	original=busy
9	yes	result=busy	original=busy
10	yes	result=busy	original=busy
11	yes	result=selective	original=selective
12	yes	result=selective	original=selective
13	yes	result=selective	original=selective
14	yes	result=selective	original=selective
15	yes	result=selective	original=selective
16	yes	result=selective	original=selective
17	yes	result=selective	original=selective
18	yes	result=selective	original=selective
19	yes	result=selective	original=selective
20	yes	result=selective	original=selective
21	yes	result=selective	original=selective
22	yes	result=selective	original=selective
23	yes	result=selective	original=selective
24	yes	result=selective	original=selective
25	yes	result=selective	original=selective

Table 66. Test results for group 3.



### 7.1.2. Labeling the data

The first task before learning is to assign labels to the tracking data gathered at the museum, shown in table 1 (chapter 3). For example, for the targeted exhibit, I need to decide whether a stop of 18 seconds should be considered ‘short’ or ‘long’. Various classification techniques can be used for this problem. Gerschenfeld [Gerschenfeld, 1999] describes a soft clustering technique using a Gaussian kernel. I have used instead hard clustering techniques, which are less precise than the soft clustering ones, as the classification problem addressed is simple enough for unsupervised classification techniques, such as k-means, to be effective.

To be able to compare results, I have actually used two different techniques for this classification problem: one is to use a popular classification procedure, known as k-means. The other consists in simply plotting histograms of the data, and finding a threshold between ‘short’ and ‘long’ stops by histogram intersection. The classes of data that are needed are only two: ‘short’ and ‘long’, as ‘skip’ is easily identifiable as a stop of zero seconds duration. I used both methods and compared results, as discussed below.

An important step before data clustering is performing data quantization, when necessary. Most statistical data analysis problems require some kind of data abstraction. The data often carries more precision than needed, and may therefore include too many categories with too much precision. This can be a problem as extra subdivisions can hide trends. Preprocessing reduces the number of variables without jeopardizing the results.

To label the data I first discarded the zero values, as the zero time length case is easily labeled as ‘skip’. From the initial 50 visitor \* 12 objects = 600 data points I obtained 413 non zero data points. I then plotted non zero data points as a histogram, such that the height of each bin corresponds to the number of data points with value  $x$ , where  $x$  is the  $x$  (horizontal) coordinate of the bin [figure 159].

Without preprocessing, the histogram has too much resolution, and its reading is ambiguous, as it shows a wave-like data envelope which hides the actual trend of the data [figure 160]. To obtain the correct data abstraction for this problem, I merged together adjacent bins, as the short bins next to the tall bins should actually be added to the tall bins to represent the actual data. I quantized the data in groups of five, and obtained the histogram shown in figure 161, which clearly exhibits an exponential decay data trend. This means roughly that visitors globally tend to make many more short than long stops at this exhibit.

I grouped the observed visitors according to their type: I observed a total of 3 greedy types, 16 busy types, and 31 selective types (table 1). These types contribute differently to the quantized histogram as shown in figures 162-164. The non exponential decay trend of the greedy type may be explained because of the small number of greedy types observed.

To distinguish short stops from long, I need to find a threshold in the quantized histogram which clearly sets the boundary between these two labels, based on the gathered tracking data. I first used the histogram intersection method and plotted normalized histograms of the greedy, busy, and selective visitor types, two at a time, as shown in figures 165-167.

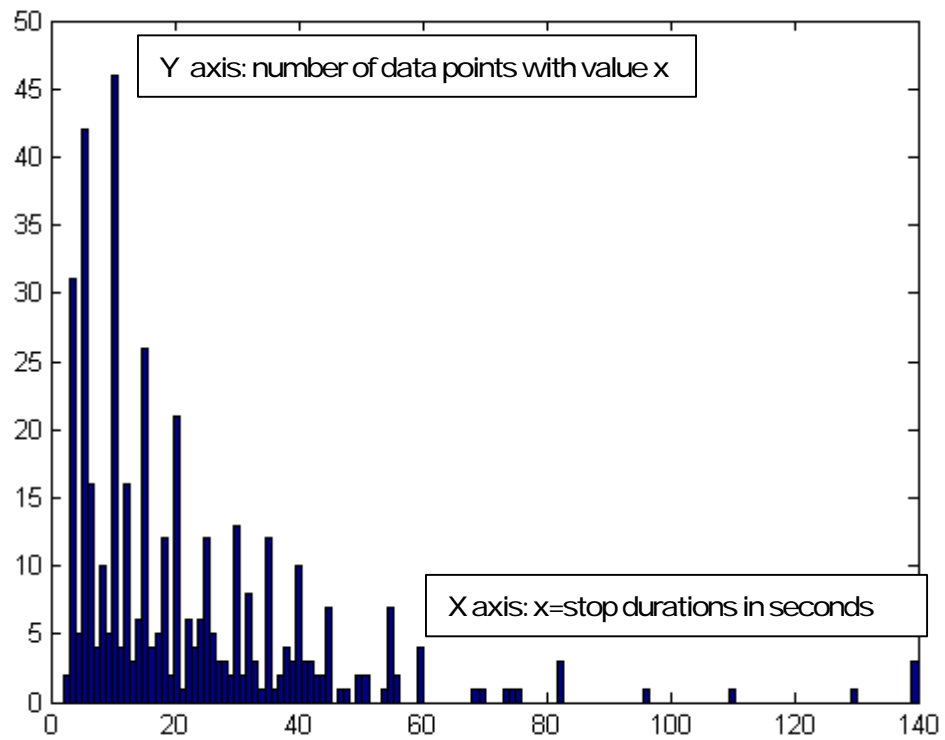


Figure 159. Histogram of the 413 non zero data points, which represent duration of visitors' stops for the 12 observed objects

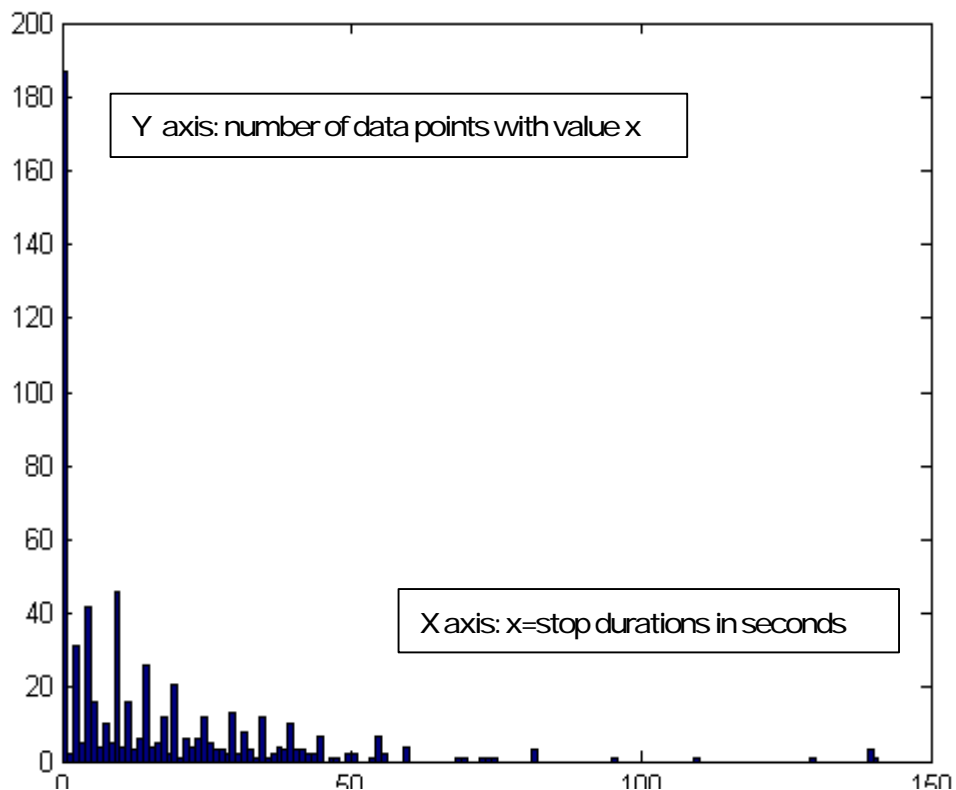
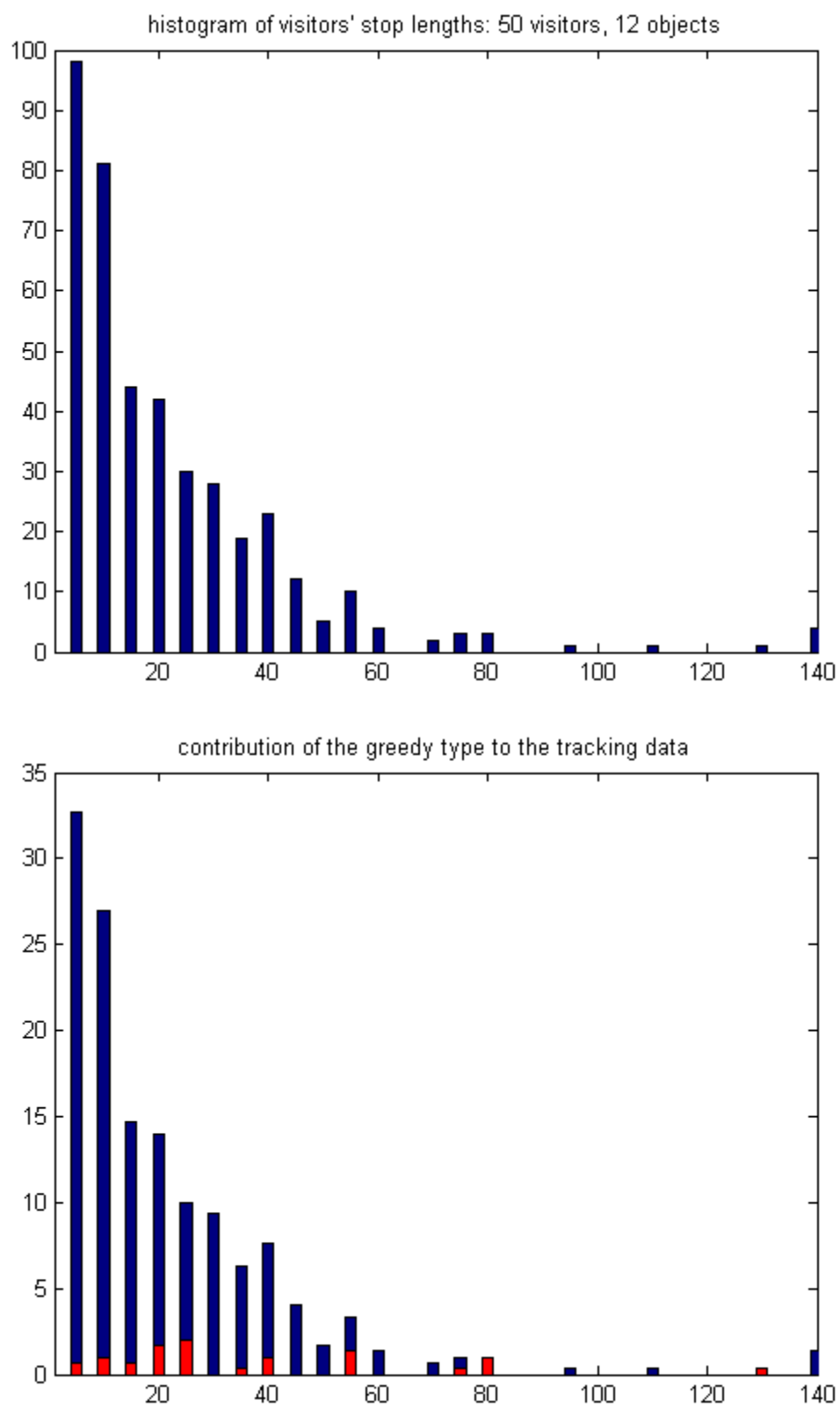
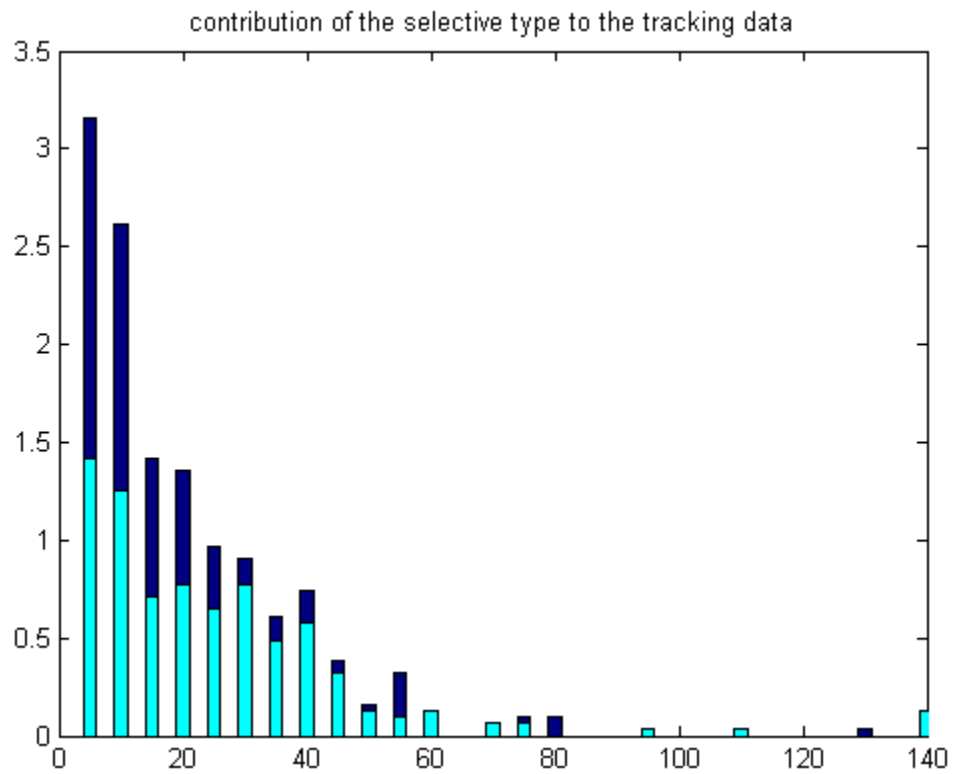
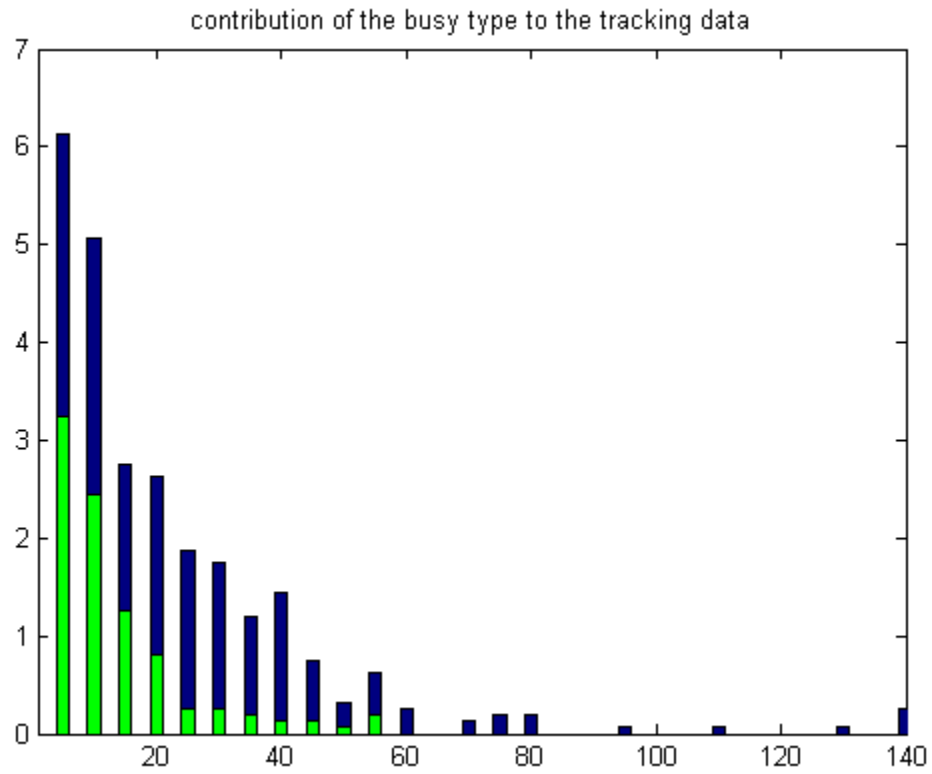


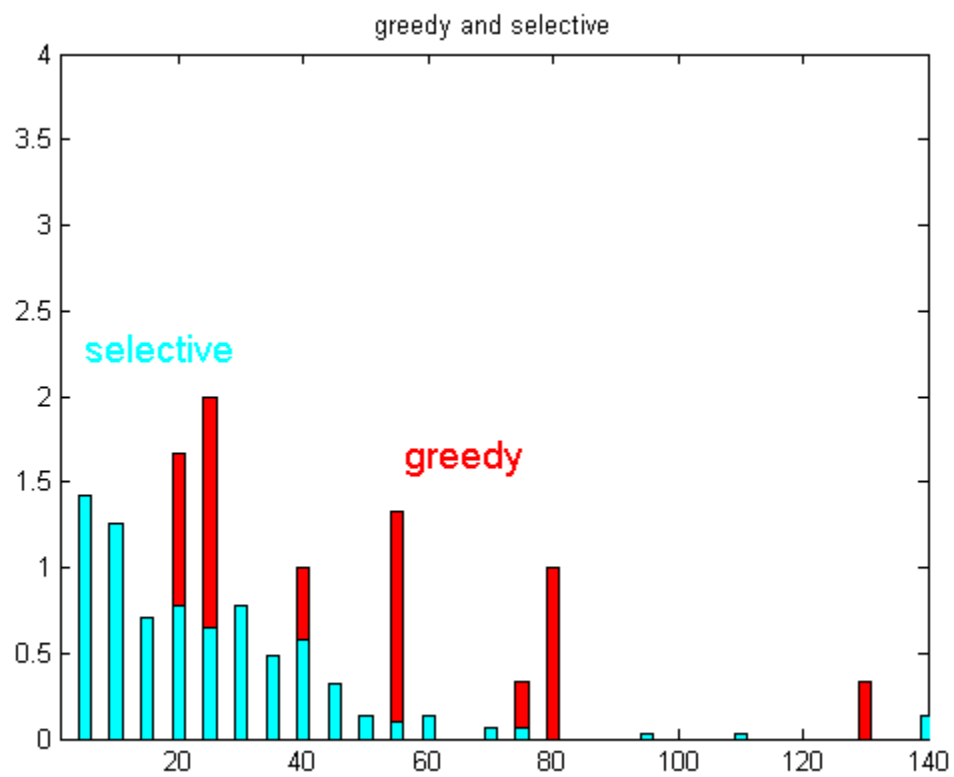
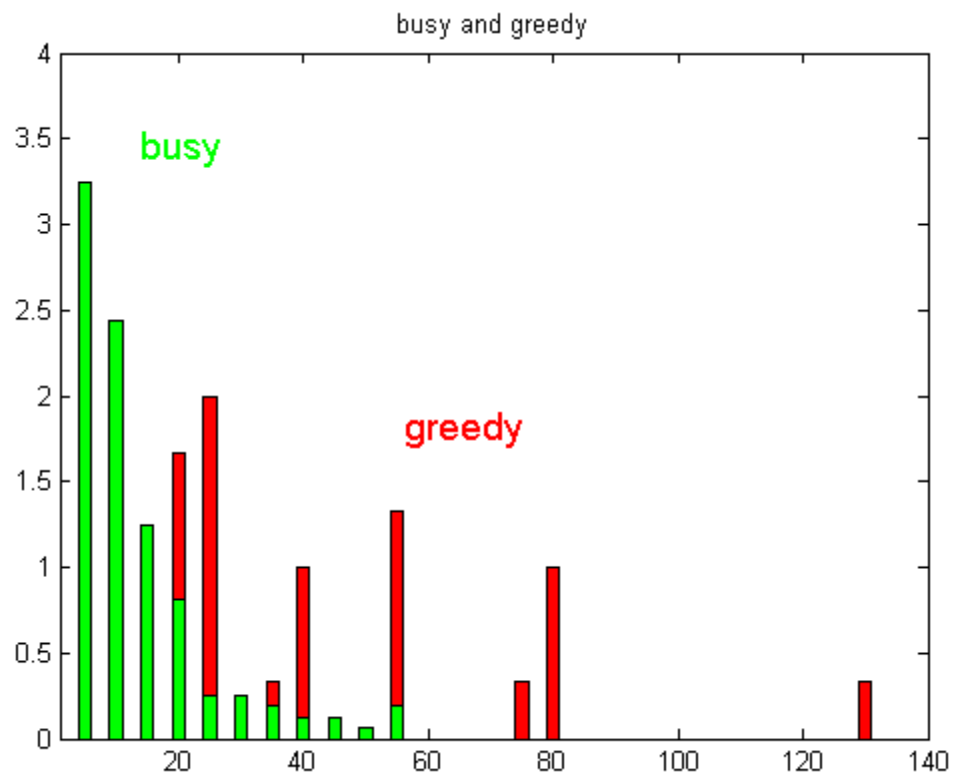
Figure 160. Histogram of the 600 data points, including the zero second length stops. The 'skip' bin is dominant.

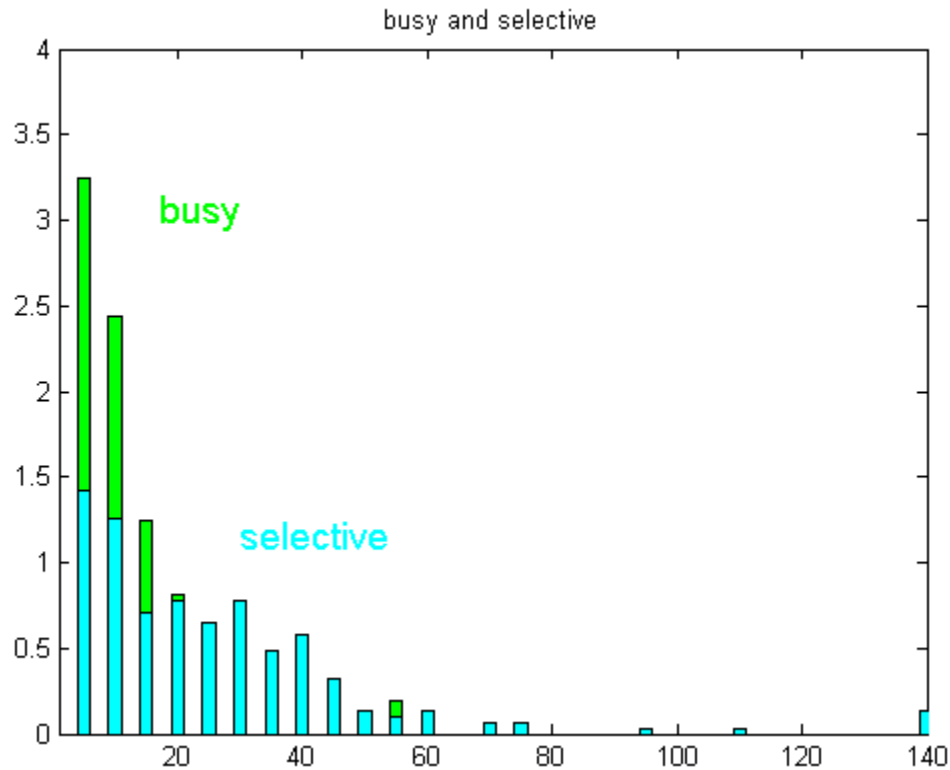


Figures 161, 162. Above. Quantized histogram for the duration of visitors' stops. Below. Contribution of the greedy type.



Figures 163, 164. Different contributions of the busy, selective (and above greedy) types to the quantized histogram of the duration of visitors' stops for the 12 observed objects.





Figures 165, 166, 167. Histogram intersection for the three selected types, taken in pairs.

It can be observed that the three histogram intersections, greedy/busy, greedy/selective, and busy/selective all overlap at bin  $x=20$ . Based on this simple analysis one can argue that a good candidate that sets the threshold between short and long stop durations for visitors at the Robots and Beyond exhibit is  $x=19$ . This means that for further data analysis, any tracked stop duration number less or equal than 19 should be labeled 'short' and any number greater than 20 should be labeled 'long'. According to this data classification, it is possible to rewrite table 1, which contains the numbers with stops durations, into a new table, which contains only three labels for the 50 visitors' stop durations at the 12 tracked museum objects on display [table 68]. Using this classification I have plotted histograms of the relative composition of 'short' and 'long' stop durations for the greedy, busy, and selective histograms [figures 171, 172, 173]

With this result it is also possible to revise the expert's original assumption on the percentage of skip, short, long stops which describe the behavior of the greedy, busy, and selective visitors respectively. These values are obtained by averaging the respective number of skip, short, long stops for the three visitor types. The new and the original parameters are similar [table 177], and the new table can be considered as a fine tuning of the parameters given by the expert.

NEW conditional probability table				original conditional probability table			
	%-skip	%-short	%-long		%-skip	%-short	%-long
Greedy	0.14	0.22	0.64	greedy	0.1	0.1	0.8
Busy	0.25	0.625	0.125	busy	0.2	0.7	0.1
Selective	0.36	0.3	0.34	Selective	0.4	0.2	0.4

Table 67. Revision of the original assumption on the percentage of skip, short, and long stops for the three targeted museum visitor types.

To confirm the threshold of 19 seconds as separator between short and long stop durations, I have also performed clustering of the data using the k-means algorithm. K-means [Therrien, 1989] is an unsupervised classification algorithm which works as follows:

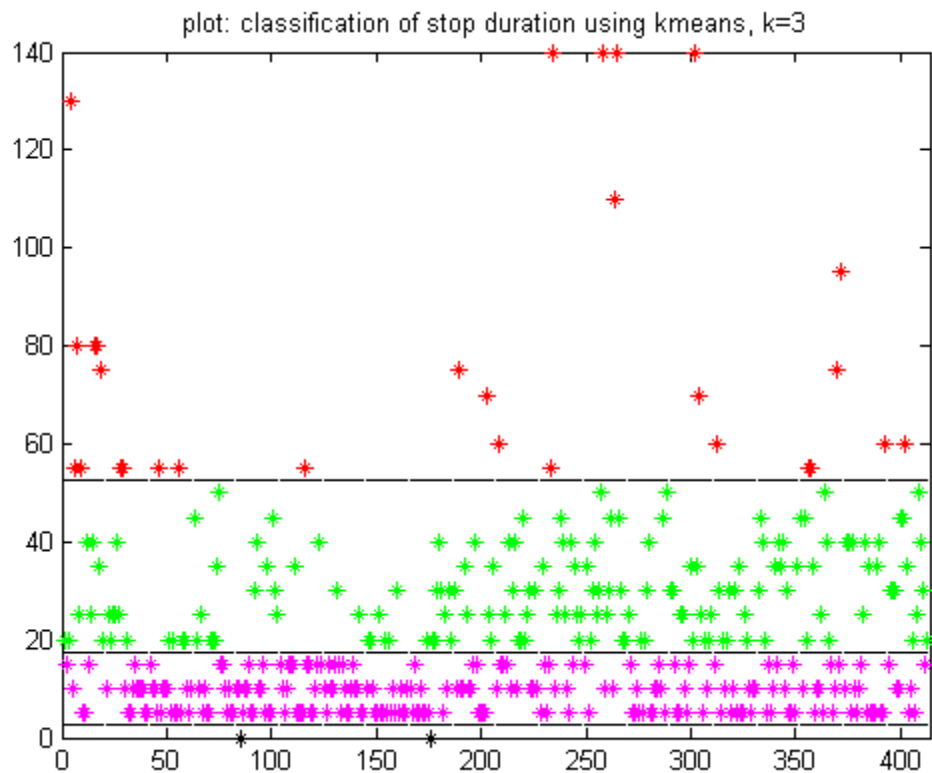
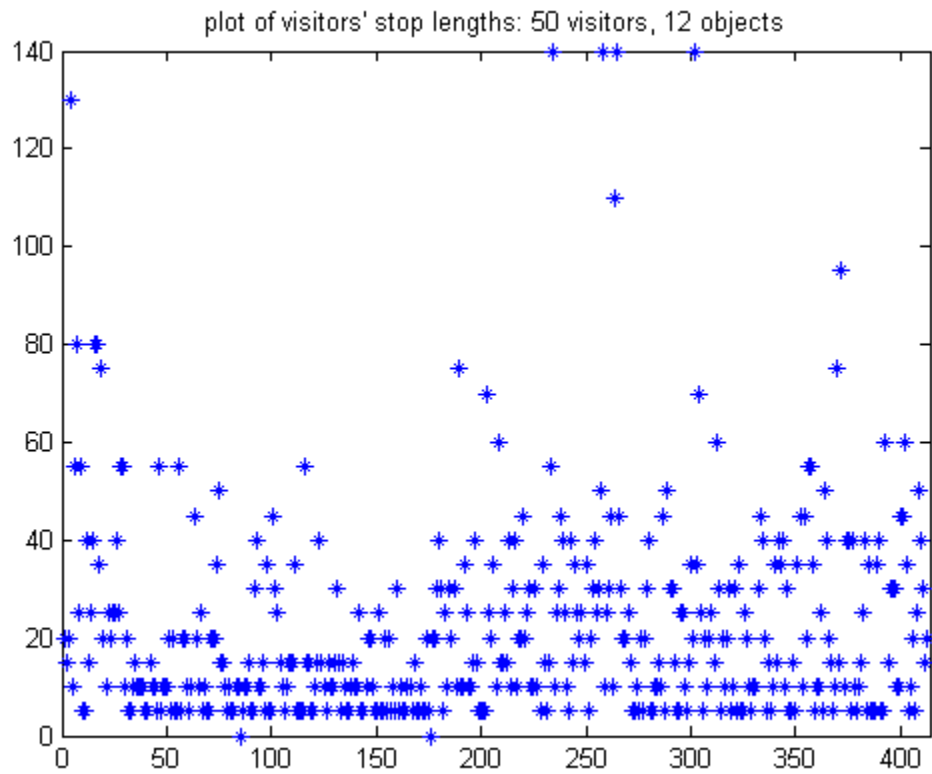
- 1. Begin with an arbitrary assignment of samples to clusters, or begin with an arbitrary set of cluster centers, and assign samples to nearest centers.
- 2. Compute the sample mean of each cluster.
- 3. Reassign each sample to the cluster with the nearest mean.
- 4. If the classification of all samples has not changed, stop.  
Else go to step 2.

Together with assigning approximate center clusters at start, the K-means algorithm requires setting  $k$ , which represents the number of clusters the algorithm divides the samples into. I performed k-means analysis of the visitor tracking data, having set  $k=3$ , to allow the extra third cluster to collect all the samples which do not fall into the 'short' or 'long' category. I also set the center cluster at start to the approximate initial values of: 10, 30, and 60 seconds respectively.



#	Intro 1	Lisp 2	MinskyArm 3	RoboArm 4	Falcon 5	Phantom 6	CogsHead 7	Quad 8	Uniuro 9	Dext Arm 10	Kismet 11	Baby Doll 12	TYPE
1	skip	skip	long	short	long	long	short	long	long	long	long	short	greedy
2	skip	short	long	short	long	long	skip	long	long	long	long	short	greedy
3	short	long	long	long	long	long	long	long	long	skip	short	long	greedy
4	skip	short	short	short	short	skip	short	skip	short	short	short	short	busy
5	short	skip	short	short	short	short	skip	skip	skip	skip	long	skip	busy
6	short	short	short	short	short	short	short	short	short	skip	long	skip	busy
7	short	short	short	short	short	short	long	short	skip	short	skip	long	busy
8	short	short	short	short	long	short	short	long	long	short	skip	skip	busy
9	short	skip	skip	skip	short	short	short	short	short	short	skip	short	busy
10	short	short	short	short	short	short	short	long	long	short	short	short	busy
11	short	long	short	short	long	long	long	short	short	short	short	short	busy
12	short	short	long	skip	skip	skip	short	short	short	short	long	short	busy
13	short	skip	skip	skip	short	skip	short	short	short	long	short	short	busy
14	short	skip	skip	skip	short	skip	skip	short	short	short	long	short	busy
15	short	short	short	short	short	short	skip	short	short	short	long	skip	busy
16	short	short	short	skip	skip	skip	short	long	short	short	skip	skip	busy
17	short	long	short	short	long	short	long	short	short	short	long	skip	busy
18	short	short	short	short	skip	short	skip	skip	skip	short	short	short	busy
19	short	short	short	skip	skip	short	short	skip	skip	skip	long	short	busy
20	short	skip	long	long	long	long	skip	skip	long	short	long	skip	sctv
21	short	long	long	long	short	skip	long	skip	skip	skip	skip	short	sctv
22	skip	skip	short	long	skip	skip	long	short	short	short	skip	long	sctv
23	short	skip	short	short	short	skip	short	skip	skip	long	long	skip	sctv
24	short	skip	long	short	long	short	skip	skip	skip	short	long	short	sctv
25	skip	long	short	long	long	short	long	skip	skip	short	skip	long	sctv
26	long	long	long	skip	skip	short	short	long	short	short	skip	skip	sctv
27	skip	skip	skip	skip	long	short	short	skip	short	long	long	short	sctv
28	long	long	long	long	long	short	short	skip	skip	skip	long	skip	sctv
29	skip	skip	short	long	long	long	skip	long	short	long	skip	short	sctv
30	skip	short	skip	skip	long	long	skip	skip	skip	long	long	long	sctv
31	long	skip	skip	skip	skip	skip	short	long	long	long	short	long	sctv
32	long	long	long	long	short	long	short	short	short	skip	short	skip	sctv
33	short	skip	skip	long	short	short	long	long	short	short	skip	skip	sctv
34	short	short	short	skip	skip	short	long	short	long	skip	long	skip	sctv
35	long	short	skip	short	skip	skip	short	long	long	short	skip	skip	sctv
36	short	short	long	long	long	long	skip	skip	long	skip	long	skip	sctv
37	short	long	short	long	skip	skip	skip	long	short	long	skip	long	sctv
38	short	long	short	long	long	short	short	long	skip	skip	skip	skip	sctv
39	short	long	short	short	long	long	skip	skip	skip	skip	skip	short	sctv
40	short	skip	short	short	skip	long	long	long	skip	long	short	skip	sctv
41	short	short	skip	short	long	short	long	skip	long	skip	long	skip	sctv
42	short	long	short	short	skip	skip	short	skip	skip	short	skip	long	sctv
43	skip	long	short	skip	long	skip	long	long	long	long	skip	skip	sctv
44	short	short	short	long	short	long	long	skip	skip	skip	long	short	sctv
45	short	skip	short	skip	long	short	long	short	long	skip	long	skip	sctv
46	short	long	short	skip	short	short	short	skip	long	skip	long	skip	sctv
47	short	skip	long	skip	short	skip	short	skip	short	long	long	short	sctv
48	short	long	short	long	long	long	short	skip	skip	skip	skip	skip	sctv
49	short	long	long	long	long	skip	skip	skip	short	short	long	skip	sctv
50	short	long	long	long	long	skip	skip	skip	short	long	skip	skip	sctv

Table 68. Labeled data



Figures 168, 169. Plot of samples (above), and color labeling of clusters after k-means classification (below).

K-means gives three clusters centered respectively in 9, 30, and 79, and with ranges 1-19, 20-54, and 55-140. I have labeled these clusters as ‘short’, ‘long’ and ‘very long’. Just as for the previous method, K-means also gives a threshold of 19 seconds as separator between short and long stops, for visitors at MIT Museum’s Robots and Beyond exhibit. These results are summarized in figure 170 below.

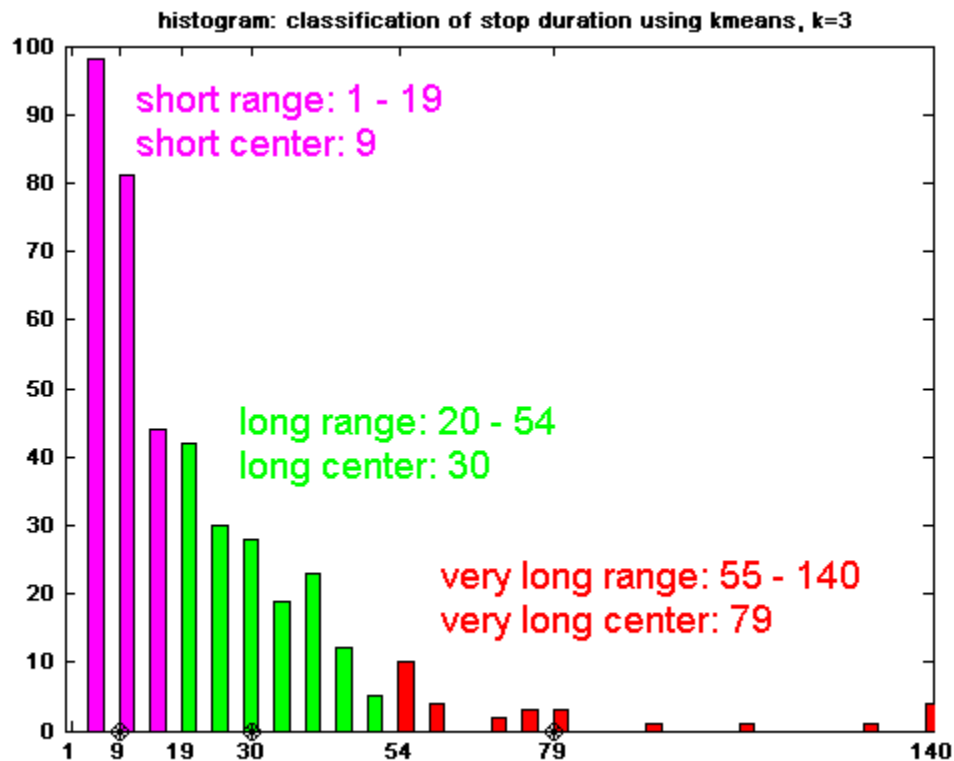
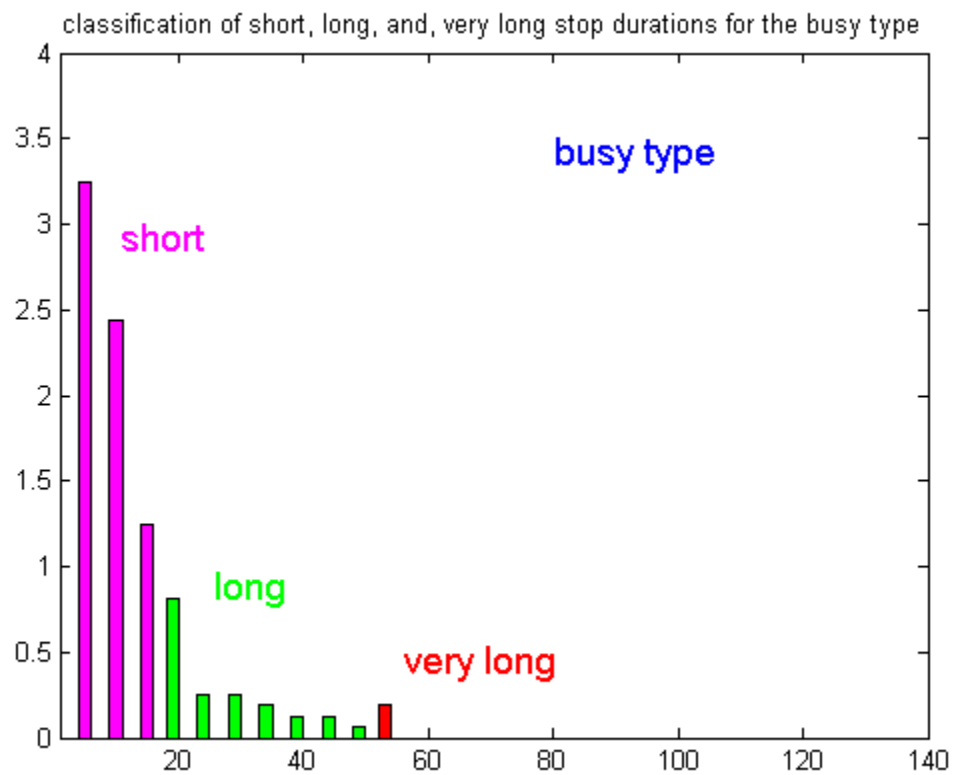
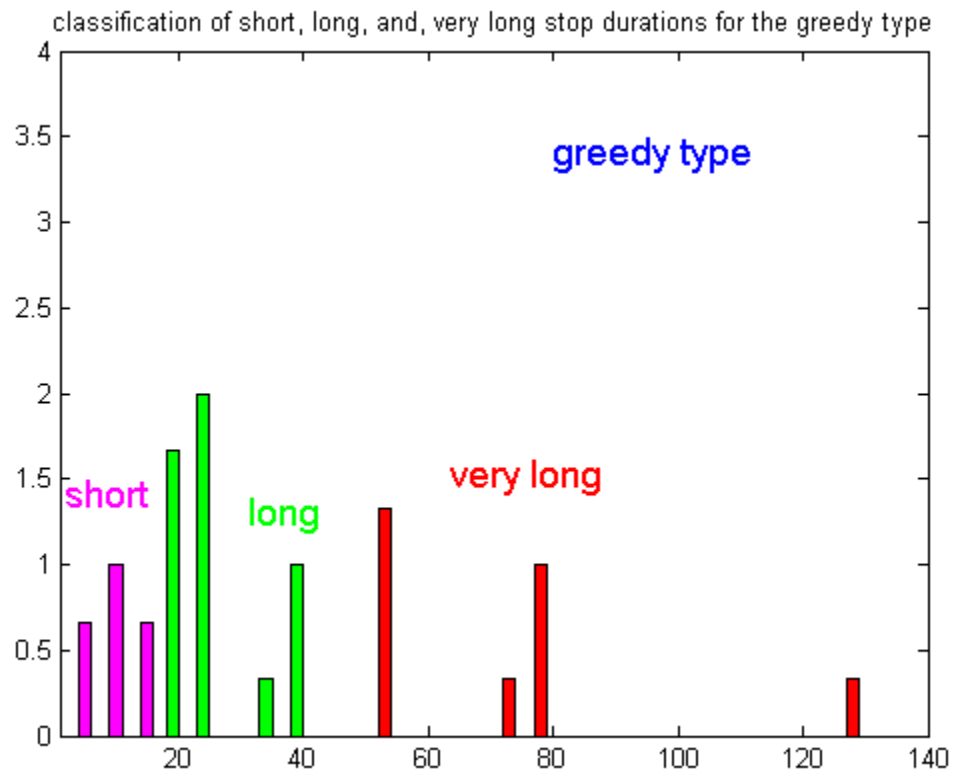
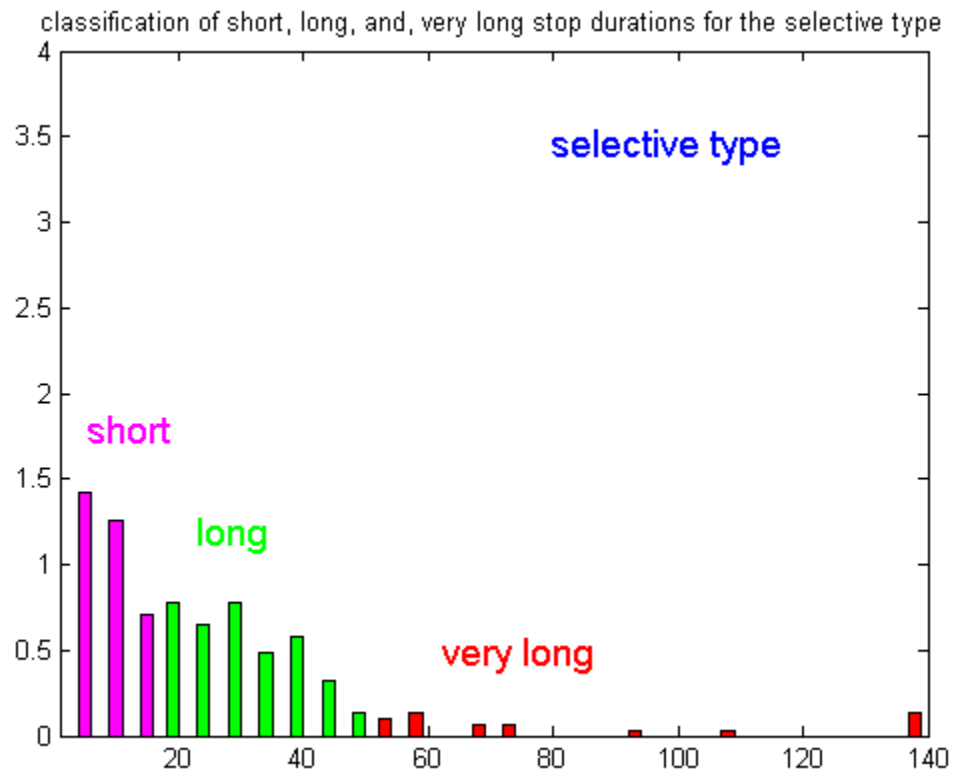


Figure 170. Summary of results after k-means classification, k=3.

With the knowledge of these ranges, I have plotted separate normalized histograms of stop durations for the three types, by color coding the bins corresponding to the different duration ranges found. For further discussion I will merge the long and very long clusters into one ‘long’ cluster, ranging from 20 to 140 seconds.





Figures 171, 172, 173. Classification of short, long, and very long stop durations for the greedy, busy and selective types.

## 7.2. Comparison with previous real-time sensor-driven content selection architectures

Learning the parameters of the system from the visitor tracking data, as shown in the previous section, is an important step in system validation, as it allows the Bayesian network to fine tune or modulate the curator's opinion about the public's typology and interests to the actual visitors' behavior. This new information is reflected in the new learned parameters of the network.

Another step in system validation is to show that the chosen approach has several advantages over other possible authoring approaches. I have illustrated in Chapter 2 a taxonomy of authoring techniques, which I have called scripted, responsive, and behavioral. This section develops a general purpose example of a typical input-output coupling problem which typically arises in interactive multimedia storytelling. The example is purposely chosen to be as abstract as possible to ensure the general validity of the following discussion. Using this example as a guideline, I will then compare authoring approaches for the specific research described in this document.

Let's imagine a sensor driven system which presents some form of digital content (video/audio/graphics/music) to the user as a function of a set of recognized input actions performed by the user. More specifically, we would have a sensing system which detects a sensorial percept, that we shall call  $P$ . This signal needs to be classified as belonging to one out of four recognized input actions:  $A_1, A_2, A_3, A_4$ . Based on which of these actions is recognized, the system selects six possible outcomes, digital content segments  $S_1, S_2, S_3, S_4, S_5, S_6$ , which are presented to the user via a computer screen, a projection, or any selected output medium. This is in essence a typical problem that multimedia authors are confronted with.

Let's analyze how we can address this problem using first a scripted approach, then a responsive approach, followed by a behavioral approach, and finally using sto(ry)chastics. For each of these cases I will then extend the conclusions of this discussion and compare approach to the one earlier described for the museum wearable application.

In the *scripted* case we typically want to encode in the system which sequence of inputs determines which output or sequence of outputs, i.e. should  $A_1, A_2, A_3, A_4$ , happening in this order cause  $S_3$ , or if  $A_3$  happens before  $A_2$  should instead  $S_4$  be shown. If the system has a memory as long as the number of possible actions, for a given sensorial percept  $P$ , there are  $4^4=256$  possible sequences which express the different order in which these actions can happen (permutations with repetition). For each of these cases we can choose an output sequence in 6 different ways, for a total of  $256 \times 6=1536$  possible cases that the author needs to keep track of. This is obviously too much to do, and therefore an obvious simplification would be to reduce the memory of the system to only the last and the current action. With this simplification there are only  $4^2=16$  possible action sequences, which determine  $16 \times 6=96$  cases: a more tractable number of possibilities to encode in the system. The problem comes when we may want to apply some changes to the original input-output coupling strategy, such as adding more output, i.e. a new content segment  $S_7$ , or adding a new input, i.e. action  $A_5$ . In either case the system modeller would have to specify what input produces which output all over again from scratch ! Therefore such an authoring technique is quite inflexible to allow the author try out various scenarios, which is usually required when initially testing and fine tuning the system's operation. If, in addition to selecting one content segment, we wish to have a criteria to assemble together and

edit a small story using as components the available outputs, as shown for the museum wearable in Chapter 6, then the cases to keep track of become:  $16 \times 6! = 16 \times 720 = 11520$ , a nearly impossible number of cases to handle. Figure 174 shows graphically what the scripted approach entails: each action node is connected to all output segments nodes.

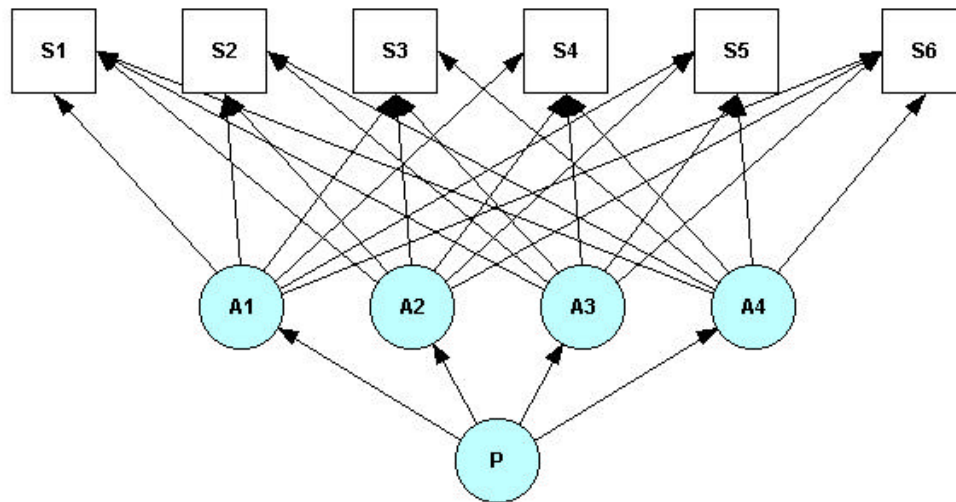


Figure 174. Scripted authoring.

If we used instead a *responsive* authoring approach, the input-output mapping problem would be quite simpler: we would establish a geography of one-to-one correspondences between inputs and outputs as shown in figure 175. As I observed in Chapter 2, this authoring strategy has an advantage not only for the author, as it simple to encode, but also for the public, as the input-output mapping is consistent in time, and it can be quickly understood or discovered by the public. It is therefore a desired approach to author several multimedia experiences. The main drawback of this technique is that it can offer the public only a shallow depth of content. If we want to add more outputs to the system we also need to add more inputs, thereby complicating the user interface to a number of actions which cannot all be memorized or learned quickly by the user. A good analogy for a responsive system is the piano, a musical instrument which associates a musical note to each key pressed. If we had a piano with only 4 keys, we would be able to produce only very limited music with it. However adding more keys complicates the interface to the extent that it can take several years and skills for the user to actually learn the interface to an extent that it produces interesting or satisfying (musical) content.

The *behavioral* approach retains some of the mapping simplicity of the responsive approach, while still allowing for some of the ability to articulate content offered by the scripted approach. In this case the system adopts a decentralized strategy to couple input and outputs, by making each output object “responsible” for triggering the appropriate event (show itself, play, stop, rewind) in correspondence to the incoming input action performed by the user. This is done in analogy with the behaviorist ethological or psychological theories which describe animal or human behavior as driven by a set of learned or innate stimulus-response associations.



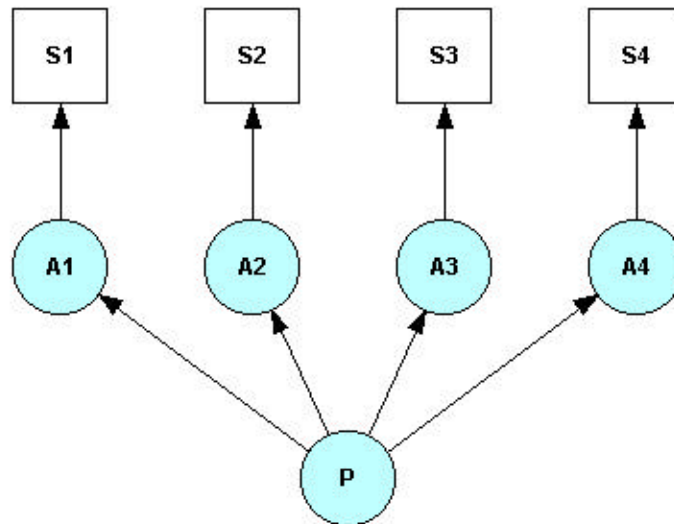


Figure 175. Responsive authoring.

The *behavioral* approach retains some of the mapping simplicity of the responsive approach, while still allowing for some of the ability to articulate content offered by the scripted approach. Typically a tree-like graphical structure is used to represent each behavior driven object. Root nodes correspond to high level behaviors which are decomposed, lower in the tree graph, into simpler behaviors. Leaf nodes represent to the atomic actions the object need to perform for the corresponding root behavior to happen. Behavior systems implement an action selection mechanism which provides a criteria to select the most appropriate behavior for the object given the set of sensory inputs present at each time step [Johnson, 1994; Blumberg, 1995]. In relation to the discussed example, we can imagine a simplified behavior system with 6 objects: S1, S2, S3, S4, S5, S6, which have one behavior only: play the corresponding video or audio clips segment. Each object/segment encapsulates an action selection mechanism which make it play (select its only behavior) as a function of one or more actions. This simplified behavior system is represented by figure 176.

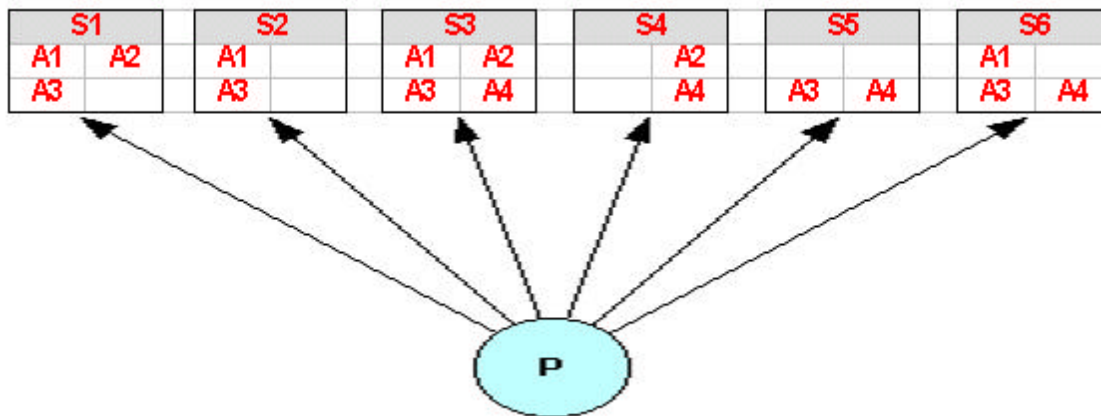
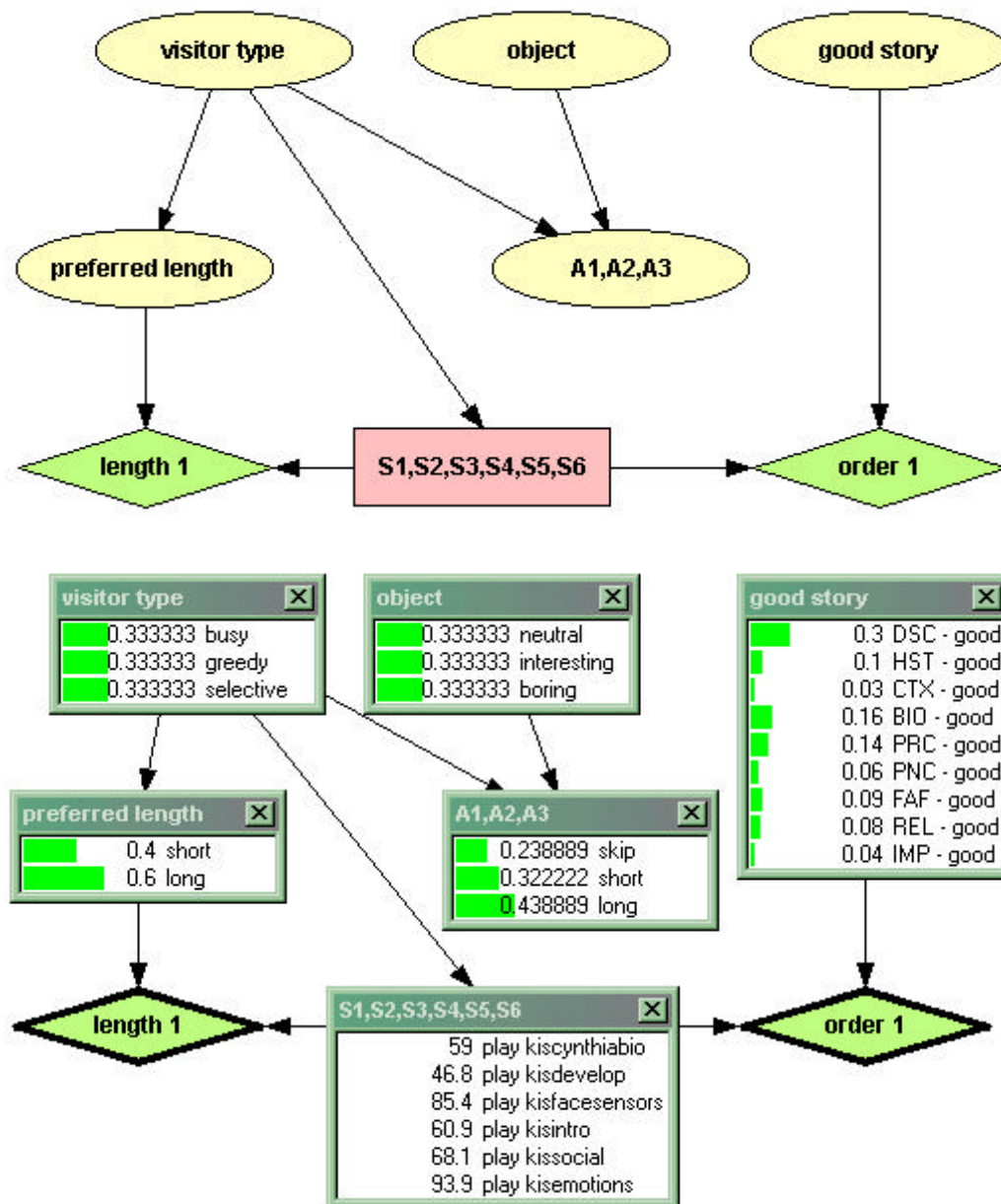


Figure 176. Behavior based authoring.

While the behavior based approach has proven effective for interactive computer graphics, where the tree-like hierarchical structure is useful to coordinate the motion of articulated graphical characters, it still leaves some open questions for the case considered in this section. For example if action A4 is recognized, objects S3, S5, S6 could all play. Yet the system would have to negotiate amongst the three activated objects, if only one of them or all should play, and in the latter case in which order.

If we limit the content selection mechanism for the museum wearable described in Chapter 6 to just one time slice (one object), we have a problem quite similar to the one illustrated above. Figures 177,178 shows the clip selection Bayesian network of the previous chapter, with renamed nodes, but same topology.



Figures 177, 178. Sto(ry)chastics authoring.

As shown from these figures, sto(ry)chastics offers several advantages over the previous approaches. With a responsive approach the resulting system would show only one pre-chosen video clip or pre-edited story per object on display. While such a system could still be used to test the public's reaction and acceptance of the museum wearable, it may overall not satisfy neither the curator, who needs to make in advance an careful selection of all the available audiovisual material, nor the public which may find the content shown either too detailed, or not enough detailed, or simply not matching their interests. With sto(ry)chastics we can aim at building technology which does not get in people's way, which does not oblige museum goers to take a pre-arranged path, and which does not require visitors to select and push buttons to get additional information from multimedia kiosks. With sto(ry)chastics a busy type is not delayed by a lengthy explanation of the artwork, and a greedy type gets all the desired details. A scripted approach would produce a system frozen in its final mapping between inputs and outputs, and would not allow the curator to easily add new objects, new content, or even new visitor types, as we can easily do with sto(ry)chastics as shown in Chapter 6. A behavior based modeling approach, while preferable to the responsive and scripted approaches, would still not allow the designer to easily derive and refine a model for the visitor type (and preferences) or tailor content to the visitor's needs. Additionally it would also not be easy to model a notion of "good story" which can easily be changed and tested with different probability priors to satisfy the curator's requirements for content presentation. Another clear advantage of sto(ry)chastics is its robustness with respect to wrong or inaccurate sensory information. Using any of the other authoring approaches, in the absence of a user model, any noise in the sensor measurements would result in a system error. If, for example, noise in the sensing system produced a short instead of a long stop duration for a greedy type, a system with a direct input-output mapping architecture would show a short introductory story according to the "misinterpreted" input signal. Sto(ry)chastics lowers instead the probability for greedy, and is still consistent with what the visitor expects: a lengthy and detailed story for the object they are looking at. If the visitor were instead to change his/her behavior at the museum, sto(ry)chastics would be able to adapt from greedy to selective or busy in a few time slices, thereby emulating a human who does not instantaneously change his/her mind about somebody's behavior, but who would use more than one observation and evidence to modify their original opinion.

By way of the posterior probabilities of the network nodes [figure 179] sto(ry)chastics gives a great deal of information: it tells about the most likely visitor's type, how interesting or boring an object may be, and which is the best segment to play taking into account the curator's criteria for a good story. It also provides robustness in evaluating this information, as explained above. It allows the curator to quickly add content, objects, and other visitor types, or even to quickly change the definition of a good story in accordance to the simplified definition of story for a computational storytelling machine given in Section 6.1. As shown in Chapter 6, the network can be extended to edit various segments pertaining to the same object together, without having the combinatorial complexity and non-flexibility of a scripted system. Sto(ry)chastics is therefore a preferable authoring technique to the scripted, responsive, or behavior-based approaches, when the content requires more depth than simple one to one mappings between inputs and outputs, when the system output needs to be personalized for individual users, and when we wish or need to model contextual or domain knowledge about the problem so as to influence the selection criteria for the output.

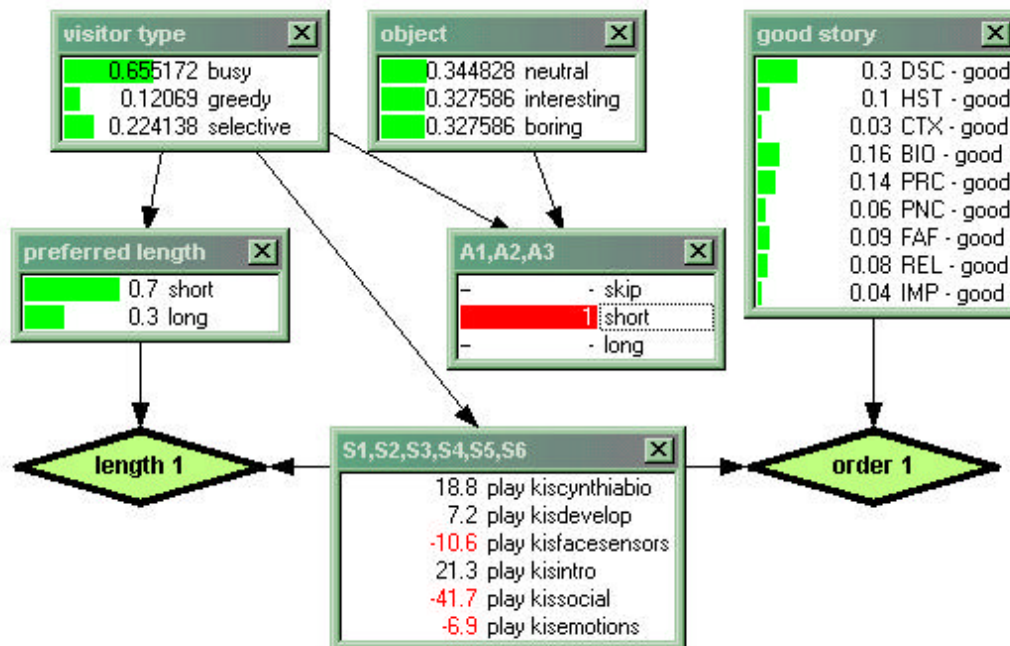


Figure 179. Richness of information provided by the sto(ry)chastics authoring.

## Chapter 8

# The impact of the museum wearable on exhibit design

The museum wearable has the potential to impact traditional exhibit design, by guiding the visitor through the exhibit, and providing him/her with information/entertainment via the wearable. The wearable could augment the visitor's experience and it expand it in ways complementary to the traditional means that the exhibit designer relies on. Some of these traditional spatial narrative aids which are used to guide and attract the visitors are space layout, text labeling, printed images, signage, or handouts. I have carried out a visualization exercise to help the reader imagine some of the possible changes in space design determined by the availability of the museum wearable to the public, specifically for the MIT Museum's Robots and Beyond exhibit. The changes that the visualization shows and proposes will need to be validated by an actual installation of the wearable at the museum, and by evaluating the implication of such installation with the exhibit designer and the curator. This visualization exercise is motivated by the fact that, because of our limited access to the exhibit's site and our limited ability to eliminate and/or move objects around, it seemed easier to imagine the impact of the museum wearable in the exhibit design as a three dimensional animation.

The digital architect of the 20th century has a broad range of tools at their disposal [Mitchell, 1993; Bendikt, 1991]. Realizing a three dimensional visualization, rather than a traditional series of sketches and drawings, allowed me to think more comprehensively about the potential impact of the museum wearable on exhibit design. I selected to use Alias Wavefront's Maya 3, for modeling and animation, so as to be able to easily move objects around or create video sequences in which certain objects would appear or disappear.

This visualization project started therefore with the production of an accurate three dimensional model of the physical layout of the exhibit space [figure 180]. It then proceeded with imaging a visitor in the space and by creating an animation of how visitors currently cruise the exhibit. To create a realistic animation, the visitor tracking data for all exhibit objects was averaged for all visitors, and proportional stop duration times were used to create the animation. The animation helps illustrate that, with the current space organization and layout, people spent most of their time at the exhibit looking at the video played by button activated kiosks, rather than with the objects on display. While the video is certainly useful and informative to explain the origin and functioning of the robots, the resulting exhibit seems to be centered around the video kiosks: the robots on display have more a decorative than a protagonist role.

The first illustration also reveals that visitors spend much of their time reading text labels related to adjacent robots. However from the tracking data and our observation of museum visitors, we have remarked that people do not spend sufficient time to read all of what is described in the posters to absorb the corresponding information. A great deal of the space occupied by the posters and text labels is therefore wasted, as most people don't take advantage of information provided in a textual form.

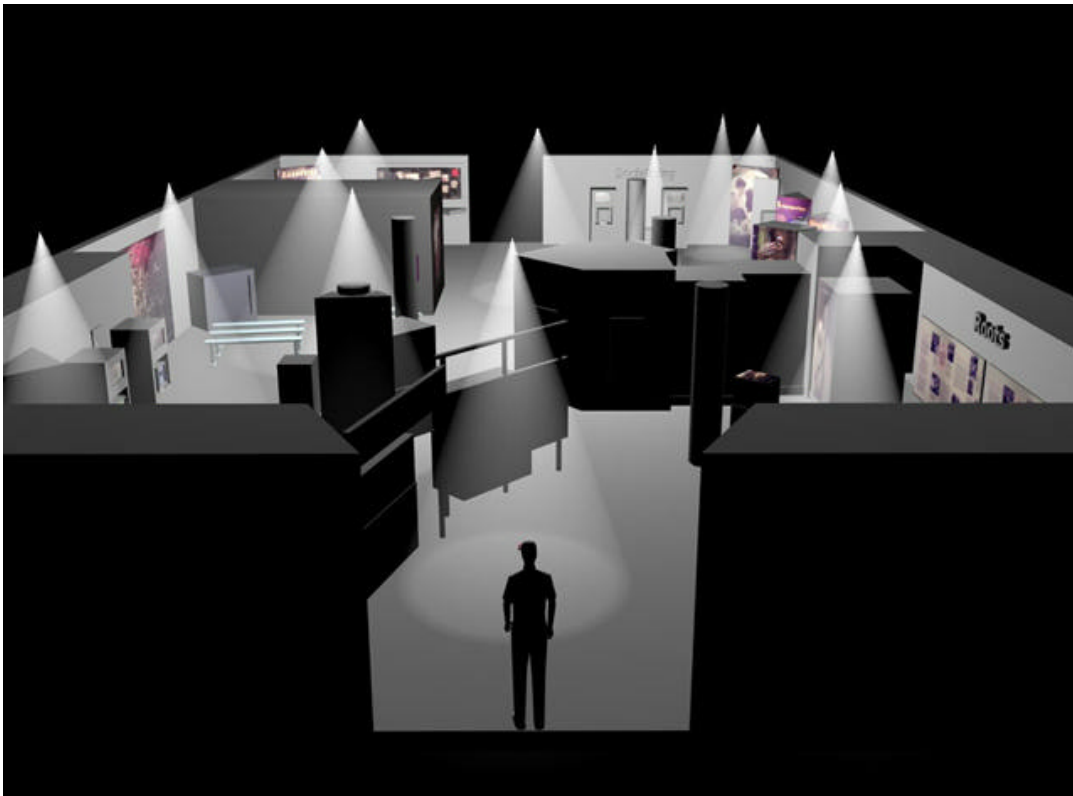
The series of animations I realized for this project help illustrate the potential changes and improvements that the wearable can produce for the space design of the MIT Museum's Robots and Beyond exhibit:

- 1. There would be no more need to have so many posters and text labels, as the corresponding information could be provided in a more appealing audio visual form, in a video documentary style by the museum wearable. The space now made available by eliminating the large posters, can be used to display more robots, which are the true protagonists of the exhibit. Typically most exhibit have to discard many interesting objects as there not enough physical space available in the museum galleries for all objects. Therefore making more space available is a clear advantage provided to the exhibit designer and the curator. Figures 181-182 show how the posters at the entrance of the MIT Museum's Robots and Beyond exhibit can be replaced by more objects to be seen and appreciated by the public.
- 2. Visitors would be better informed, as the information currently provided by the posters is mostly neglected by the public. The same information would instead become part of the overall narration provided by the wearable, and it would be better absorbed and appreciated by the public.
- 3. The video kiosks would no longer be necessary because the same material would be presented by the museum wearable. Therefore the robots would be again the center of attention for visitors, as the wearable's display allows both the real world and the augmented audiovisual information to be seen at the same time as part of the wearer's real surround view. This would again make more space available for additional objects to be displayed.

The fact that the audio visual material is presented together with the corresponding object by the museum wearable, rather than separately in space and time, in a museum catalogue or in a printed poster, or in a video or multimedia kiosk, is also of great importance. While no studies have been conducted yet on the quality and effectiveness of the learning experience offered by the museum wearable, there is reasonable hope believing that synchronous and local information provided while actually looking at the object described by the wearable can make a longer and more effective impression on the visitor.

One last series of animations shows a futuristic imaginary *Empty Space* [figure 183], with no objects, which can be used by a museum to show *any* past exhibits, possibly with the

aid of a three dimensional holographic head mounted display. In such a space, people could select which past exhibit they wish to see, using a dial attached to the museum wearable, and multiple people could be in the same space exploring different exhibits which have been hosted in that space in the past.

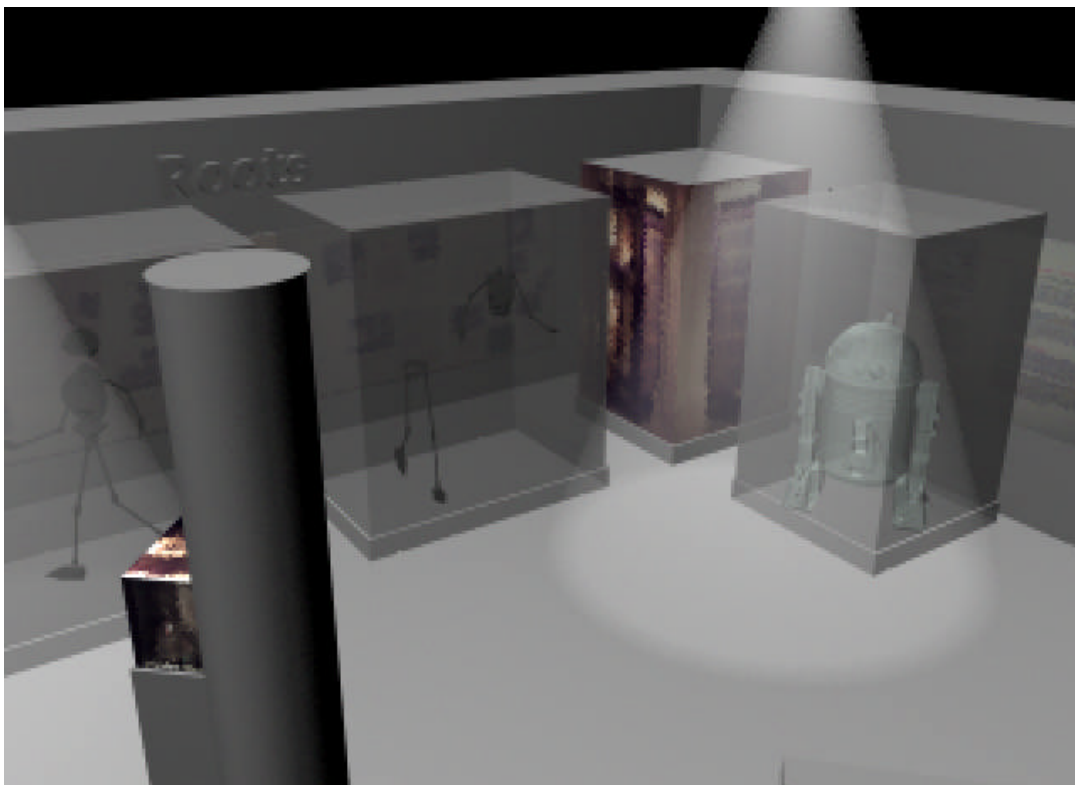


**Figures 180. 3D model: start frame of animation of a visitor at the exhibit.**





Figures 181. Before use of the museum wearable: posters in the Roots section are needed to explain the exhibit.



Figures 182. Potential impact of the museum wearable on the current exhibit layout: the posters in the Roots section are replaced by new objects.

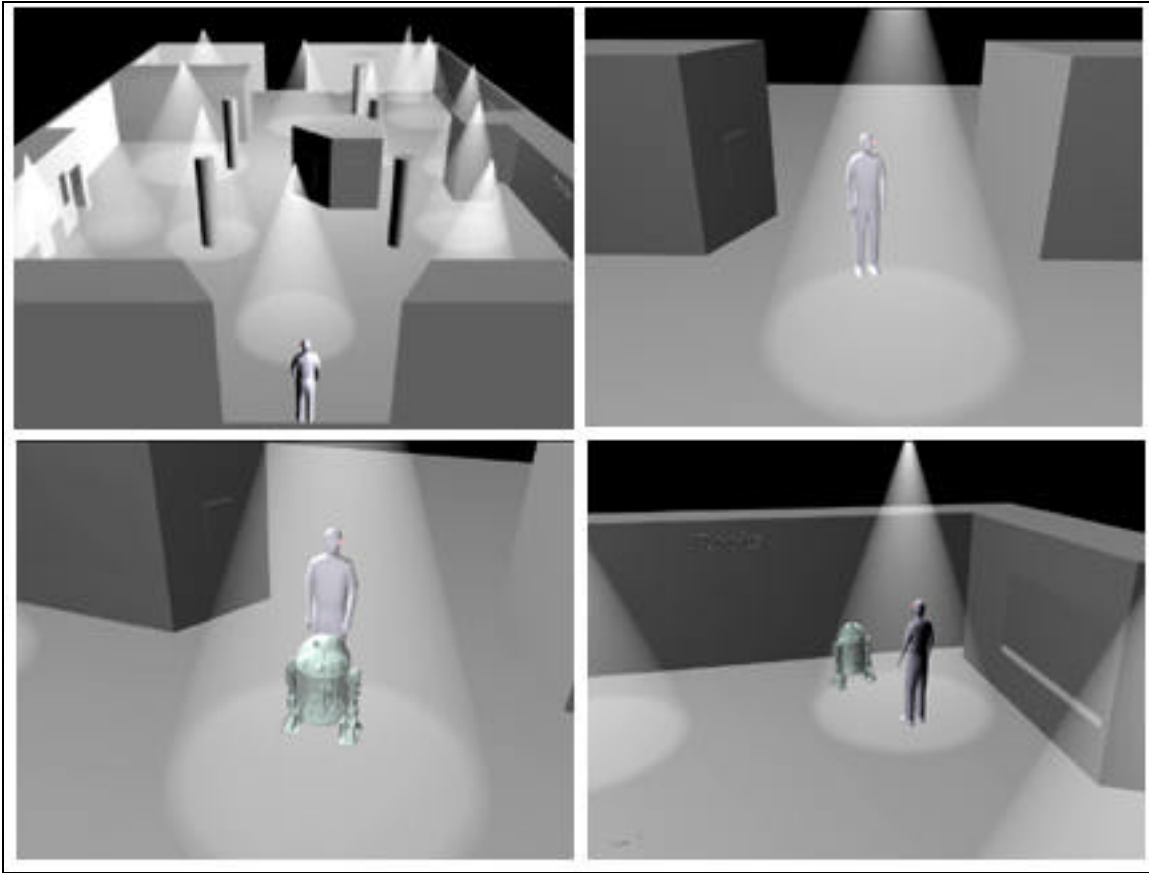


Figure 183. Visualization of a futuristic *Empty Space*: no objects are physically present in the space.

An empty space, just with walls and the infrared location sensors, can be used by a museum to show any past exhibits, possibly with the aid of a three dimensional holographic head mounted display. In such a space, people could select which past exhibit they wish to see, using a dial attached to the museum wearable. Multiple people could be in the same space exploring different exhibits which have been hosted in that space in the past. The frames of the above animation show an overview of the Empty Space, a visitor approaching an area covered by the location sensor, and then an object appearing to the visitor as he/she enters the area.



## Chapter 9

# Summary of Accomplishments and Future Directions

This research introduces Bayesian networks for real time sensor-driven storytelling, and demonstrates that they are a useful tool to model the uncertainty in the sensor measurements, make informed guesses about people's intentions during interaction, encapsulate the storyteller's message, and orchestrate a complex audiovisual narration as a function of these. I call such stochastic modeling of story and user-story interaction: sto(ry)chastics. Sto(ry)chastics has implications both for the human author (designer/curator) which is given a flexible modeling tool to organize, select, and deliver the story material, as well as the audience, which receives personalized content only when and where it is appropriate.

Sto(ry)chastics proposes an alternative to complex centralized interactive entertainment programs which simply read sensor inputs and map them to actions on the screen. Interactive storytelling with such one-to-one mappings leads to complicated control programs which have to do an accounting of all the available content, where it is located on the display, and what needs to happen when/if/unless. These systems rigidly define the interaction modality with the public, as a consequence of their internal architecture. Rather than directly mapping inputs to outputs, we need to endow digital content itself with the ability to "understand the user" and to produce an output based on the interpretation of the user's intention in context.

Sto(ry)chastics uses dynamic Bayesian networks to model the sensors in the system and allows the system to interpret the sensor data by taking into account the context and domain of interaction, represented by other nodes of the network. The interpretation of sensor data is robust in the sense that it is probabilistically weighted by the history of interaction of the participant as well as the nodes which represent context. Therefore noisy sensor data, triggered for example by external or unpredictable sources, is not likely to cause the system to produce a response which does not "make sense" to the user. For content selection and delivery, sto(ry)chastics allows the system to build a profile of the participant through time, and therefore can tailor content according to the participant's estimated desires and interests. These features: robustness with respect to "misunderstandings" because of knowledge of context, and the ability to learn more about the user through time, produce a system which with further development can potentially, in the future, simulate an elementary conversation with a human participant and is able to gear the topic of discussion towards the interests of the latter.

As an example of application of sto(ry)chastics, and to illustrate its features, I have designed and developed a real time storytelling device: a museum guide which in real time evaluates the visitor's preferences by observing his/her path and length of stops along the museum's exhibit space, and selects content from a set of available movie clips, audio, and animations. This device, which I call the Museum Wearable, illustrates the advantages of sto(ry)chastics in designing and authoring real-time sensor-driven digital media presentation systems.

With further testing on site, the museum wearable will possibly enrich and personalizes the visit as a visual and auditory storyteller which can adapt its story to the audience's interests and guide the public through the path of the exhibit. With this device curators may be able to present a larger variety and more connected material in an engaging manner within the limited physical space available for the exhibit.

The museum wearable identifies three visitor types: busy, greedy, and selective, which have been selected as the essential museum visitor types from the museum literature. It uses a custom-made infrared location sensor to gather tracking information about the visitor's path in the museum's gallery and uses this information to introduce evidence in the dynamic Bayesian network which interprets the sensor information and delivers content to the visitor. The network performs probabilistic reasoning under uncertainty in real time to identify the visitor's type. It then delivers an audiovisual narration to the visitor as a function of the estimated type, interactively in time and space. The model has been tested and validated on observed visitor tracking data using the EM algorithm. Estimation of visitor preferences using additional sensors is provided in a simulated environment.

The main contribution of this research is to show that (dynamic) Bayesian networks are a powerful modeling technique to couple inputs to outputs for real time sensor-driven multimedia audiovisual stories, such those that are triggered by the body in motion in a sensor-instrumented interactive narrative space. Other contributions are: the design of the museum wearable application, the assembly and fashioning of a wearable computer, specifically conceived for museum use; the design and realization of a new long range infrared location identification sensor; the construction and test of a variety of Bayesian networks for user type and profile estimation; the extension of the previous Bayesian network for real time story segment selection and editing; model selection; model validation and parameter learning via the EM algorithm; and simulation of processing multiple sensor inputs with a Bayesian network for robust estimation and more accurate user profiling.

More specifically, Chapter 4 and Sections 5.1. and 5.2. have illustrated a variety of Bayesian networks for sensor-driven user modeling and user-driven content selection, in the framework of the museum wearable application. I have called this type of modeling: sto(ry)chastics. Through these examples the reader can observe several virtues and advantages of sto(ry)chastics, which I list and summarize below. Sto(ry)chastics is:

- 1. *Flexible*: it is possible to easily test many different scenarios by changing the parameters of the system, performing probability update, and reading the posterior

probabilities of the parameters. These changes include adding and removing states for a node, changing the prior probabilities for the root nodes, or changing the conditional probability tables for some nodes, if the posterior probabilities do not model the problem in a satisfying way. The system is also flexible as it can adapt to the visitor's changing interests and curiosity.

- 2. *Reconfigurable*: it is also quite easy to add or remove nodes and/or edges from the network without having to “start all over again” and specify again all the parameters of the network from scratch. This is a considerable and important advantage with respect to hard coded or heuristic approaches to user modeling and content selection. Only the parameters of the new nodes and the nodes corresponding to the new links need to be given. The system is extensible story-wise and sensor-wise. These two properties: flexibility and ease of model reconfiguration allow the system modeller, the content designer, and the exhibit curator to work together and easily and cheaply try out various solutions and possibilities until they converge to a model which satisfies all the requirements and constraints for their project. A network can also rapidly be reconfigured for another exhibit.
- 3. *Robust*: Probabilistic modeling allows the system to achieve robustness, as the system draws conclusions (posterior probabilities) by weighting all the network parameters which in turn describe user, sensors, and story segments, probabilistically. A good example is one which assigns a probability to the visitor being excited when the GSR sensor measures various peaks at its output. This probability is set to 0.9, as our assumption is that in 10% of cases, people may be excited for other reasons than what there are seeing or listening to with the wearable. The effects of this conditional probability not being 1 but a lesser value, cascade down to other dependent parameters during the probability update operation, and therefore allow us to come up with reasonable and robust guesses about the values of the nodes of interest. Robustness also means that rather than having sensors being threshold activated, which is a strategy prone to errors simply because the real world and real sensors are noisy, the information provided by the sensors produces an action only with the given probability and if it contributes to the overall understanding of the user desires or intentions. Also, as shown in section 5.3.3. with additional sensors, and using redundant sensing modalities, sto(ry)chastics can potentially achieve robustness by sensor fusion with the addition of a few nodes to the network.
- 4. *Adaptive*: sto(ry)chastics is adaptive in two ways: it adapts both to individual users and to the ensemble of visitors of a particular exhibit. For individuals, even if the visitor exhibits an initial “greedy” behavior, by consistently making long stop durations at the exhibit objects, it can later adapt to the visitor's change of behavior if he/she starts making only short stops. It will initially guess that the visitor is possibly selective, and if the busy stops continue, it will finally label the visitor as busy. It is important to notice that, reasonably and appropriately, the system “changes its mind” about the user type with some inertia: i.e. it will initially lower the probability for a greedy type until other types gain probability. Sto(ry)chastics can also adapt to the collective body of its users. If a count of busy/greedy/selective visitors is kept for the exhibit, these numbers can become priors for subsequent visitors, thereby causing the

entire exhibit to adapt to the collective body of its users through time. This feature can be seen as “collective intelligence” for an exhibit which can adapt not just to the individual visitors but also to the set of its visitors.

- 5. *Context-sensitive*: for any system to be robust and to provide relevant information to its user, it is important to model the context of interaction together with the other system parameters. For example, using any of the modeling techniques described in the previous sections, sto(ry)chastics could provide an explanation of a visitor’s change of behavior at the museum. If suddenly a greedy type starts making short stops, the system, before concluding that they are actually a selective or busy type, could test if the current time is near closing time for the museum galleries, or if, by use of other room sensors, there is a great crowd in the new galleries where the visitor is making short stops. Coming up with the right conclusions, given this type of external information means that the system is context-sensitive. Specifically, with respect to the networks previously described, context is modeled in the object nodes that have a neutral/interesting/boring discrete state, and in the “good story” node which models the curator’s preferences as context for content selection for the targeted exhibit. Context is therefore also modeled by the priors of some root nodes. Through these priors we incorporate domain knowledge and expectations of the curators of the visitor types that a particular exhibit is likely to attract. ) Also in some cases the network topology reflects domain knowledge about the problem/application. Hence sto(ry)chastics is not just data driven but data and expert driven. In this respect it stands mid way between traditional Artificial Intelligence top down high level reasoning approaches (expert systems) and pattern recognition based bottom up low-level approaches (HMMs), and is able to include both within the same graphical probabilistic framework.
- 6. *Able to explain its choices*: as opposed to neural networks, all the nodes of a Bayesian networks have a meaning and a role, and therefore by reading the posterior probabilities for the nodes of interest, including nodes that are not observable as physical measurements, we can “understand” how the system comes up with its conclusions and how it makes its choices. What we human call “conclusions” are the result of a probability update in a Bayesian network, as explained by the examples in Section 4.1. The diagrams shown in Chapters 4 and 5 can be seen as a brain probe into the system during interaction.
- 7. *Accessible*: graphical models have a very intuitive meaning which facilitates understanding and collaboration between the curator and the technologist (when these are different people). They provide an easy-to-understand representation of conditional independence relationships for the non-mathematician and are therefore accessible to all people who wish to contribute to the design of the interactive experience modeled by sto(ry)chastics.

Finally, the accomplished research described in this document is highly interdisciplinary. Specifically, in completing my thesis I have used knowledge acquired in mathematical



probabilistic modeling, machine learning, computer programming in C++, DirectX, 3D modeling and animation, networking, electronics for the design and construction of the IR location sensors, machining for the wearable computer and head mounted display assembly, video and film editing, photography, and architecture.

An experimentation phase at the museum should follow the current research. A procedure would have to be set to establish if and how the museum wearable does actually enhance learning and entertainment at an exhibit, or how the content shown does actually match the visitor's preferences. A selected group of visitors could be instructed to "speak their mind" onto a microphone attached to the wearable during the visit, to later compare how the content shown actually corresponds to the visitor's interests. A questionnaire should also be handed to the visitor at the end of the experience. It should contain questions asking how people think the wearable has been useful or entertaining for them, as well as asking visitors specifically to give a subjective evaluation of their own interest profile, after the visit.

The research here presented can be expanded in various ways. One direction of work is to find ways to get to know the visitor better so as to target content presentation more accurately towards his/her level of knowledge or competence. Asking to the visitor to fill lengthy questionnaires upon entering the exhibit may not be practical. An interesting venue for future work may come from extracting as much information as possible from the visitor's home page in the WWW, and derive from that a set of prior probabilities for the visitor's interest profile node. With this technique the only extra time required to the visitor would be to give the system the URL address of their home page, if they wish. It is of course desirable to experiment with additional sensors, such as the GSR and the camera sensor, whose functioning and contribution to the museum wearable experience are described in Chapter 5. More visitor tracking data would need to be gathered at the museum site, to eventually infer more visitor types than the ones described in this document, and compare them with the more sophisticated visitor typologies discussed in the museum literature.

The museum wearable can also potentially be developed to become a useful tool for visitor tracking data gathering in the museum. For example, rather than coming up with a set of visitor types from the museum literature, one could use the opposite approach of "inferring" the visitor types from a statistical analysis of the tracking data (path and stop duration) gathered by visitors with the museum wearable. This information would help the curator, exhibit designer, and the modeler of the interactive museum experience to refine their knowledge about visitor types for a specific exhibit.

Similarly, by analyzing the posterior values of the object nodes, the curator and the exhibit designer could see which objects are the most interesting or boring for the visitors, and change the exhibit layout accordingly.

An important extension to the museum wearable would allow it to support visitors who want to come to the museum as a group and have the freedom to comment and discuss the artwork amongst themselves instead of being fully immersed in the experience

offered by the wearable. A simple modification to the current prototype would be to add a small microphone capable of detecting when the visitor is talking, which would automatically pause the narration. If for example the group of visitors is composed by high school students, it would be useful, for learning purposes, to make the visitor profile available to the users at the end of the exhibit, and have the system regroup the visitors according to matching profiles. This same capability could also be made available to visitors at the end of their tour in the museum's cafeteria, to play matchmaking among those who wish to be involved. Alternatively the visitor's profile, path, and length of stay can be used to create a web based exhibit catalogue whose URL can be sent to the visitor as a personalized basis for further learning.

I would also like to consider an extension of the museum wearable to my previous work on wearable city, and envision applications which involve the urban environment as a narrative space, for art, tourism, and entertainment. Another possible development is to model "media actors" [Sparacino, 1999b] to be modeled by sto(ry)chastics. Media Actors are active content which interact with the user, as a theatrical or improvisational performer would relate with its public. For example if the visitor skips an important artwork which matches his/her profile, the media actor would call them back by playing an audio clip saying "hey come back, come see me, I am really interesting, allow me to tell you my story". This venue is encouraged by personal discussions I've had with some museum curators and theater director which welcome the opportunity to create experimentation platforms to merge their fields.

In a more distant future it would also be desirable that museum exhibits using similar Bayesian networks are able to exchange experience about their visitors so that they can better adapt to their needs. They could then learn from each other to fine tune their parameters for more effective content presentation. This is a quite substantial step ahead which would require sequential learning (as opposed to the batch learning technique illustrated in Section 7.1.) not from data, but from another network. In addition a metrics to evaluate network similarity would have to be researched.

# Bibliography

- Albrecht, D.W., Zukerman, I., Nicholson, A.E., and Bud, A. "Towards a bayesian model for keyhole plan recognition in large domains." In Jameson, A.; Paris, C.; and Tasso, C., eds., *Proceedings of the Sixth International Conference on User Modeling (UM '97)*, pp. 365-376. Springer, 1997
- Andersen, S. "HUGIN – a Shell for Building Bayesian Belief Universes for Expert Systems." In: *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, IJCAI*, 1080-1085. Menlo Park, California, 1989.
- Ballard D. and Brown C. *Computer Vision*, Prentice Hall, Englewood Cliffs, NJ, 1997.
- Bates J. et al. "An Architecture for Action, Emotion, and Social Behavior." In: *Proceedings of the Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World*, 1992.
- Behringer, R., Klinker, G., Mizell, D.W. *Augmented Reality: Placing Artificial Objects in Real Scenes*. Proceedings of IWAR 98, A.K. Peters LTD, Natick, MA, 1999.
- Benedikt, M. *Cyberspace: First Steps*. The MIT Press. Cambridge, Massachusetts. 1991
- Bilmes J. A "Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models" ICSI-TR-97-021, 1997.
- Bishop C.M., *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.
- Blumberg, B., Galyean, T., "Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments." *Computer Graphics SIGGRAPH '95 Proceedings*, 30(3):47-54, 1995.
- Brand, M., Oliver, N., and Pentland, A. Coupled hidden Markov models for complex action recognition, *CVPR*, pages 994-999, Puerto Rico , 1997.
- Brand, M. "Coupled hidden Markov models for modeling interacting processes." Submitted to: *Neural Computation*, November 1996.
- Bruner J. *Acts of Meaning*, Cambridge, Ma.: Harvard University Press, 1990.
- Bruner J. *Actual Minds, Possible Worlds*. Cambridge, Ma.: Harvard University Press, 1986.
- Bruner J. *On Knowing: Essays for the Left Hand*, Cambridge, Ma.: Harvard University Press, 1962.
- Burke K. *Grammar of Motives*. University Chicago Press, 1969.
- Catania, A.C. and Harnad, S. eds., *The Selection of behavior: the operant behaviorism of B.F. Skinner: comments and consequences*, Cambridge University Press, 1988.
- Chickering D.M., Heckerman D., and Meek C. A "Bayesian Approach to Learning Bayesian Networks with Local Structure." Microsoft Research Corporation, Technical Report MSR-TR-97-07, August 1997.
- Chickering, D. M., and Heckerman, D., "Efficient approximations for the marginal likelihood of Bayesian networks with hidden variables." Technical Report MSR-TR-96-08, Microsoft Research, March, 1996.
- Conati, C., A. Gertner, K. VanLehn, and M. Druzdzel. "On-Line Student Modeling for Coached Problem Solving Using Bayesian Networks." In *UM97, 6th International Conference on User Modeling*. 1997. Chia Laguna, Sardinia, Italy, 1997.

- Cooper, G. and Herskovits, E. "A Bayesian method for the induction of probabilistic networks from data." Technical Report KSL-91-02, Knowledge Systems Laboratory, Medical Computer Science, Stanford University, 1991.
- Cooper, G.F. "Probabilistic Inference Using Belief Networks is NP-hard." Technical Report, KSL-87-27, Medical Computer Science Group, Stanford University, 1987.
- Cowell et al. *Probabilistic Networks and Expert Systems*. Springer-Verlag, NY, NY, 1999.
- Cowell, R. "Introduction to inference for Bayesian networks." In Jordan (1999), pp 9-26, 1999.
- Darrell T., Maes P., Blumberg B., and Pentland A., "A Novel Environment for Situated Vision and Behavior" in Landy, M., Maloney, L., Pavel, M., eds., *Exploratory Vision: The Active Eye*, Springer-Verlag, 1995.
- Darrell T., Moghaddam B., Pentland A., "Active face tracking and pose estimation in an interactive room." In *CVPR 96*. IEEE Computer Society, 1996.
- Davenport, G., Agamanolis, S., Barry, B., Bradley, B., and Brooks K. "Synergistic storyscapes and constructionist cinematic sharing." *IBM Systems Journal*, Vol. 39, Nos. 3 & 4, 2000b, pp 456-469.
- Davenport, G., Aguierre Smith, T., Pincever, N. "Cinematic Primitives for Multimedia". *IEEE Computer Graphics and Animation special issue on multimedia*, July, pp. 67-74, 1991.
- Dean D. *Museum Exhibition: Theory and Practice*. London, Routledge, 1994.
- Dempster, A.P., Laird, N.M., and Rubin, D.B. "Maximul Likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistocal Society*, B 39 (1), 1-38, 1977.
- Duda R.O. Hart P.E. and. Stork D.G. *Pattern Classification*. 2nd edition, Wiley, New York, NY 2000.
- Emering L., Boulic R., Thalmann D., "Multi-Level Modeling and Recognition of Human Actions Involving Full Body Motion" *Proc. Autonomous Agents 97*, ACM Press, 1997.
- Feiner, S., MacIntyre, B., Hollerer, T., and Webster, A. "A Touring Machine: Prototyping 3D Mobile Augmented Reality Systems for Exploring the Urban Environment." In: *Proceedings of the International Symposium on Wearable Computers*, pp. 74-83. Los Alamitos, CA, USA: IEEE Computer Society, 1997.
- Forbes J., Huang T., Kanazawa K., and Russell S., "The BATmobile: Towards a Bayesian automated taxi" In: *Proceedings of the Int'l Joint Conf. on Artificial Intelligence* Vol. 14, pp. 1878--1885, 1995.
- Frey, B.J. *Graphical Models for Machine Learning and Digital Communication*. MIT Press, Cambridge, 1998.
- Friedman, N. "Learning belief networks in the presence of missing values and hidden variables." In: J. Douglas H. Fisher, editor, *Machine Learning: Proceedings of the Fourteenth International Conference (ICML '97)*, pp 125-133, Nashville, TN, Morgan Kaufmann, 1997.
- Friedman, N. "The Bayesian structural EM algorithm." In: *Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI, 1998.
- German. S., and German, D. "Stochastic Relaxation, Gibbs distributions, and the Bayesian restoration of images." *IEEE Trans Pattern Analysis and Machine Intelligence*, 6, pp 721-741, 1984.
- Gerschenfeld, N. *The Nature of Mathematical Modeling*, Cambridge University Press, 1999. See pp. 178-185
- Ghahramani and Hinton, "Parameter Estimation for LDS", tech report., 1996.

- Ghahramani, Z. "Learning dynamic Bayesian networks." In: *Adaptive processing of temporal information Lecture notes in artificial intelligence*, C.L. Giles and M. Gori editors, New York, NY, Springer-Verlag, 1997.
- Gilks, W., Thomas, A., Spiegelhalter, D. "A language and a program for complex Bayesian modeling." *The Statistician* 43, 169-178, 1994.
- Hall D.L. and Llinas J. "An Introduction to Multisensor Data Fusion". In: *Proceedings of the IEEE, special issue on Data Fusion*, January 1997.
- Hayes-Roth, B. and van Gent, R. "Improvisational puppets, actors, and avatars". In: *Proceedings of the Computer Game Developers' Conference*, Santa Clara, CA, 1996.
- Healey, J., Seger, J., and Picard R. "Quantifying Driver Stress: Developing a System for Collecting and Processing Bio-Metric Signals in Natural Situations." In: *Proceedings of the Rocky Mountain Bio-Engineering Symposium*, April 16-18 1999.
- Healey, J. "Wearable and Automotive Systems for the Recognition of Affect from Physiology." MIT PhD Thesis - Electrical Engineering and Computer Science Department, June 2000.
- Heckerman, D. Probabilistic Similarity Networks, Technical Report, STAN-CS-1316, Depts. of Computer Science and Medicine, Stanford University, 1990.
- Heckerman, D. "A tutorial on learning with Bayesian networks." In Jordan, M. I. editor, *Learning in Graphical Models*. The MIT Press, Cambridge, MA, 1999.
- Heckerman, D. "A Tutorial on Learning with Bayesian Networks." In Jordan (1999), pp 301-354, 1999.
- Henze, N., and Nejd, W. "Bayesian modeling for adaptive hypermedia systems." In ABIS 99, 7. GI-Workshop Adaptivitat und Benutzermodellierung in interaktiven Softwaresystemen, Magdeburg, Sept. 1999.
- Hinckley, K., Pierce, J., Sinclair, M. and Horvitz, E. "Sensing techniques for mobile interaction." *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology (UIST 2000)*, pp. 91-100. New York, NY: ACM Press, 2000.
- Hoeting J.A., Madigan, D. Raftery, A.E., Volinsky, C.T. Bayesian "Model Averaging: A Tutorial." *Statistical Science*, Vol. 14. No. 4, pp 382-417, 1999.
- Hooper-Greenhill E. *Museums and their visitors*, London, Routledge, 1999.
- Houbart, G. *Viewpoints on Demand: Tailoring the Presentation of Opinions in Video*. MIT Masters Thesis, 1994
- Howard, R.A., and Matheson, J. E. "Influence Diagrams." In: *Applications of Decision Analysis*, volume 2, eds. R.A. Howard and J.E. Matheson, 721-762, 1981.
- Jameson, A. "Numerical uncertainty management in user and student modeling: An overview of systems and issues." In: *User Modeling and User-Adapted Interaction*, 5:193--251, 1996.
- Jebara, T., and Pentland, A. "Action reaction learning: Analysis and synthesis of human behaviour." *IEEE Workshop on The Interpretation of Visual Motion*, 1998.
- Jelinek F. *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, 1997.
- Jensen, F.V. *An Introduction to Bayesian Networks.*, UCL Press, 1996.
- Jensen, F.V. *Bayesian Networks and Decision Graphs*. Springer-Verlag, New York, 2001.
- Jensen, F.V., Lauritzen, S.L, and Olesen, K.G. "Bayesian updating in causal probabilistic networks by local computations." *Computational Statistics Quarterly* 4, 269-282, 1990.

- Johnson M., WavesWorld: PhD Thesis, *A Testbed for Three Dimensional Semi-Autonomous Animated Characters*, MIT, 1994.
- Jordan M.I., editor. *Learning in Graphical Models*. The MIT Press, 1999.
- Jordan M.I., Ghahramani Z., Jaakkola T.S., Saul L.K. "An introduction to variational methods for graphical models", In Jordan (1999), pp 105-161, 1999.
- Kaelbling L.P., Littman M.L., and Cassandra A. "Planning and Acting in Partially Observable Stochastic Domains". *Artificial Intelligence*, Vol. 101, 1998.
- Kim, J.H. and Pearl, J. "A computational model for causal and diagnostic reasoning in inference systems." In: A. Bundy ed. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, Karlsruhe, Germany, pp 190-193, Morgan Kaufmann, 1983.
- Kinderman, R. and Snell J.L. "Markov Random Fields and Their Applications." *American Mathematical Society*, Providence, USA, 1980.
- Klein, L. *Exhibits: Planning and Design*. Madison Square Press, New York, pp70-71, 1986.
- Koller, D. and Pfeffer, A. "Probabilistic Frame-Based Systems." In: *Proceedings of the Fifteenth National Conference on Artificial Intelligence*; Madison, Wisconsin; July 1998.
- Kwon J. and Murphy K. "Modeling Freeway Traffic with Coupled HMMs". Submitted to *NIPS*, 2000.
- Lauritzen, S.L. and Spiegelhalter D.J. "Local computations with probabilities on graphical structures and their application to expert systems." *Journal of the Royal Statistical Society, Series B* 50, 157-224, 1988.
- Lauritzen, S.L. "The EM algorithm for graphical association models with missing data." *Computational Statistics and Data Analysis* 19, pp 191-201, 1995.
- MacKay, D.J.C. (1992a) "Bayesian interpolation." *Neural Computation*, 4(3):415--447, May 1992.
- Mackay, D.J.C. (1992b) "The evidence framework applied to classification networks." *Neural Computation* 4 (5), pp 720-736, 1992.
- MacKay, D.J.C. "Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks." *Network: Computation in Neural Systems* 6: pp 469-505, 1995.
- Madsen, A.L. and Jensen, F.V. "Lazy propagation: A junction tree inference algorithm based on lazy evaluation." *Artificial Intelligence* 113, 203-245, 1999.
- Magnenat Thalmann N., Thalmann D. "The Artificial Life of Synthetic Actors" *IEICE Transactions*, Japan, invited paper, Vol. J76-D-II, No. 8, pp. 1506-1514, August 1993.
- Matthew Brand, Nuria Oliver, and Alex Pentland. "Coupled hidden markov models for complex action recognition." In: *Proceedings of IEEE CVPR97*, 1996.
- Michael I. Jordan, editor, *Learning in Graphical Models*, Kluwer Academic Publishers, Dordrecht, 1998.
- Minka T. "From Hidden Markov Models to Linear Dynamical Systems". MIT Tech Report, 1999.
- Minka, T. *A family of algorithms for approximate Bayesian inference* PhD thesis, MIT EECS, January 2001.
- Mitchell T.M., *Machine learning*, Mc Graw-Hill, New York, 1997.
- Mitchell, W.J. Virtual Architecture: Computers are Challenging Notions of Space. *Architecture*, 82(12), p. 39, 41, 43, December, 1993.

- Murphy K. "A Survey of POMDP Solution Techniques". Internal Report. Berkeley, September 2000.
- Murphy, K. and Mian S. "Modeling gene expression data using dynamic Bayesian networks." Technical report, Computer Science Division, University of California, Berkeley, CA, 1999.
- Murphy, K. "The Bayes Net Toolbox for Matlab" In: *Computing Science and Statistics: Proceedings of Interface*, volume 33, 2001.
- Neal, R.M. "Probabilistic Inference Using Markov Chain Monte Carlo Methods." Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, September 1993.
- Neapolitan, E. *Probabilistic Reasoning in Expert Systems*. John Wiley and Sons, New York, 1990.
- Oliver N., Pentland A., Berard F., LAFTER: "A real-time Lips and Face Tracker with Facial Expression recognition." *CVPR 97, IEEE Computer Society*, San Juan, Puerto Rico, June 1997.
- Oliver N., Rosario B, and Pentland A. "A Bayesian Computer Vision System for Modeling Human Interactions". In: *Proceedings of Intl. Conference on Vision Systems ICVS99*. Gran Canaria. Spain. January 1999.
- Paradiso J.A. "The Brain Opera Technology: New Instruments and Gestural Sensors for Musical Interaction and Performance." *Journal of New Music Research*, 28(2), pp 130-149, 1999.
- Pasula, H., Russell, S. "Approximate Inference for First-order Probabilistic Languages." In: *Proc. IJCAI-01*, Seattle, 2001.
- Pavlovic V., Rehag J., Cham T-J., Murphy K. "A Dynamic Bayesian Network Approach to Figure Tracking Using Learned Dynamic Models". *Proceedings of ICCV (Int'l Conf. on Computer Vision)* 1999.
- Pavlovic, V.I. *Dynamic Bayesian Networks for Information Fusion with Applications to Human-Computer Interfaces*. PhD Thesis. University of Illinois at Urbana-Champaign, 1999.
- Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.
- Perlin, K., Goldberg, A. "Improv: A system for scripting interactive actors in virtual worlds." In: *Computer Graphics, SIGGRAPH 96 Proceedings*, pp 205-216, ACM, 1996.
- Pfeffer, A., Koller, D., Milch, B., Takusagawa, K.T. "SPOOK: A system for probabilistic object-oriented knowledge representation." In: *Proc. UAI*, 1999.
- Pynadath D. V. and Wellman M. P. "Accounting for context in plan recognition, with application to traffic monitoring." *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, 1995, pp. 472-481, 1995.
- Qi, Y., Minka T., Picard R.W. "Automatic Determination of Relevant Features for the Bayes Point Machine." unpublished, 2001.
- Rabiner L., Juang B-H. *Fundamentals of Speech Recognition*. Prentice Hall, 1993.
- Saul, L.K. and Jordan, M.I. "Exploring tractable substructures in intractable networks." In D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo editors, *Advances in Neural Information Processing Systems* 8, MIT Press, 1996.
- Sawhney N., Balcom D., Smith I. "Authoring and Navigating Video in Space and Time: An Approach Towards Hypervideo." *IEEE Multimedia*, Vol. 4, No.4, October-December 1997.
- Sawyer R.K. *Creativity in Performance*, Ablex, 1997a.
- Sawyer R.K. *Pretend Play as Improvisation*, Erlbaum, 1997b.

- Schiele, B., Oliver, N., Jebara, T., and Pentland, A. "An Interactive Computer Vision System, DyPERS: Dynamic and Personal Enhanced Reality System." *International Conference on Computer Vision Systems*, CVPR, Gran Canaria, Spain, Springer, pp. 51-65, 1999
- Schödl A., Essa I. "Machine Learning for Video-Based Rendering". *NIPS* 2000.
- Schödl A., Szelinski R., Salesin D., Essa I. "Video Textures". *SIGGRAPH* 2000.
- Serrell, B. "The question of visitor styles." In S. Bitgood (Ed.). *Visitor Studies: Theory, Research, and Practice*, Vol. 7.1. (pp. 48-53). Jacksonville AL: Visitor Studies Association, 1996.
- Shafer, G., and Shenoy, P. "Probability propagation." *Annals of Mathematics and Artificial Intelligence* 2, 327-352, 1990.
- Smyth, P. "Belief Networks, Hidden Markov Models, and Markov Random Fields: A Unifying View." *Pattern Recognition Letters*, 1998.
- Sparacino F., Davenport G., Pentland A. "Media in Performance." *IBM Systems Journal*, Vol. 39, Nos. 3 & 4, 2000b, pp 479-510.
- Sparacino F., Davenport G., Pentland A., "Wearable Cinema/Wearable City: bridging physical and virtual spaces through wearable computing." *IMAGINA 2000*, Montecarlo, January 31st-Feb 3, 2000a.
- Sparacino F., Larson K., MacNeil R., Davenport G., Pentland A. "Technologies and methods for interactive exhibit design: from wireless object and body tracking to wearable computers." In: *Proceedings of the International Conference on Hypertext and Interactive Museums*, ICHIM 99, Washington, DC, Sept. 22-26, 1999a.
- Sparacino F., Davenport G., Pentland A., "Media Actors: Characters in Search of an Author." *IEEE Multimedia Systems '99, International Conference on Multimedia Computing and Systems* (IEEE ICMCS'99), Centro Affari, Firenze, Italy 7-11 June 1999b.
- Sparacino, F. et al., "City of News." In: *Proceedings of Ars Electronica Festival* 1997, Linz, Austria, 8-13 September 1997.
- Sprites P., Glymour C., and Scheines R. *Causation, Prediction, and Search*. MIT Press, 2000.
- Starner, T., and Pentland, A. "Visual Recognition of American Sign Language Using Hidden Markov Models." *International Workshop on Automatic Face and Gesture Recognition* (IWAAGR). Zurich, Switzerland, 1995.
- Starner, T., Mann, S., Rhodes, B., Levine J., Healey, J., Kirsch, D., Picard, R., and Pentland A., "Augmented Reality through Wearable Computing." *Presence*, Vol. 6, No. 4, pp. 386-398, August 1997.
- Terzopoulos D. "Artificial life for computer graphics." *Communications of the ACM*, 42(8), 32-42, August, 1999.
- Terzopoulos, D., Tu, X., Grzeszczuk, R. "Artificial fishes: Autonomous locomotion, perception, behavior, and learning in a simulated physical world." *Artificial Life*, 1, 4, pp. 327-351, December, 1994.
- Therrien, C.W. *Decision, Estimation, and Classification*. John Wiley and Sons, 1989.
- Tosa, N. and Nakatsu, R. "Life-like Communication Agent – Emotion Sensing Character 'MIC' and Feeling Session Character 'MUSE'". *Proceedings of the International Conference on Multimedia Computing and Systems*, pp.12-19, 1996.
- Vygotsky L.S. "Imagination and Creativity in Childhood". *Soviet Psychology*, 1990, 28.1., pp 84-96.
- Weiss Y., "Belief propagation and revision in networks with loops." MIT AI Memo 1616 (CBCL Paper 155). Presented in NIPS\*97 workshop on graphical models, 1997.



Welch G., Bishop G. "An Introduction to the Kalman Filter". University of North Carolina at Chapel Hill, Department of Computer Science, TR 95-041, 1999.

Whittaker, J. *Graphical Models in Applied Multivariate Statistics*, John Wiley and Sons, 1990.

Wolpert, D.H. "On the use of evidence in neural networks" in C.L. Giles, S.J. Hanson, and J.D. Cowan editors, *Advances in Neural Information Processing Systems 5*, pp 539-546, Morgan Kaufman, 1993.

Wren C., Azarbayeani A., Darrell T., Pentland A., "Pfinder: Real-time tracking of the human body". *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):780-785, July 1997.

Wren C., Sparacino F., et al., "Perceptive spaces for performance and entertainment: Untethered interaction using computer vision and audition." *Applied Artificial Intelligence*, 11(4):267-284, June 1997a.

Yedidia J.S., Freeman W.T. and Weiss Y. "Generalized Belief Propagation." *NIPS* 2000.

Zweig, G.G. *Speech Recognition with Dynamic Bayesian Networks*. PhD Thesis, University of California, Berkeley, 1998.

## Acknowledgments

I would like to thank my thesis committee: Neil Gershenfeld, Walter Bender, Kent Larson, and Glorianna Davenport. I would also like to thank Ron MacNeil for offering thesis advice. Various experts have provided precious scientific advice for this work: Alex Pentland, Jim Rehg, and Bill Freeman. Special thanks to my mentors: Yuan Qi and Tom Minka: their teaching has been at the core of the motivation and developments of this research. I also received additional scientific and thesis advice from Martin Szummer and Ali Azarbayejani.

I would also like to acknowledge Kevin Murphy, for his BTN library and the Hugin developers for the Hugin Bayesian Net library.

Without the help of my undergraduate UROPers collaborators I would not have been able to develop this research as fast and well. It's also been fun to work with them and they've often exceeded my expectations, working late hours and surprising me in many ways. Their names are, in alphabetical order: Anjali D'Oza, Eric Hilton, Sarah Mendelowitz, Audrey Roy, Chin Yan Wong.

I would like to express my special thanks to all the MIT Museum people who have joined me in this venture, and whose help and collaboration has been indispensable for this research: director of exhibitions Beryl Rosenthal, former director of exhibitions Janis Sacco, and MIT Museum director Janet Pickering.

My deepest thanks to Dean Staton, and Tony to Robinson, from MIT's President Office.

Many staff members at the Media Lab have helped in many ways to make this research happen. Some of them are: Linda Peterson, Susan Bottari, and Linda Lowe.

My thoughts go to Patrick Purcell who had a role in encouraging me to apply to the MIT Media Lab to carry out the interdisciplinary work I wanted to do, and who has been a great mentor and friend in my years at MIT.

My parents have provided infinite support at all times. Without them I would not have been able to do all that I have done so far.

Last but not least, my many friends in this area have provided a great source of support and inspiration in many ways. I have personally thanked them for their care.

## Appendix

# Sto(ry)chastics for other applications

Sto(ry)chastics is suited for a variety of applications, typically those in which the space or the user are instrumented with sensors, and a strategy is needed to couple sensory media inputs with a coordinated presentation of story fragments, which together form a story in the sense defined in Chapters 1 and 6. I illustrate in this section another possible application of sto(ry)chastics, based on my previous work, and sketch out how I would approach the real time input-outputs coupling problem with a Bayesian network approach. The purpose of this section is to provide the reader with a developed example of the sto(ry)chastics authoring approach, so as to show how the research presented in this document can extend to other interactive art and entertainment applications.

For this purpose I selected to present the Tabletop circus, a sophisticated interactive application for which interpreting the participant's intentions to couple inputs to outputs seems essential to achieve the artist's goal. In Tabletop Circus I envision an entertaining home-theater interactive show in which participants interact gesturally with projected images of circus performers on a custom-made table-size stage [figures 184, 185]. The work builds on principles incorporated into Unbuilt Ruins, an interactive museum design installed in the Compton Gallery at MIT in February 1999. Inspired by Alexander Calder's Circus (on permanent exhibit at the Whitney Museum in New York City) Tabletop Circus will invite the participant to use their hand to spin acrobats, to push a tightrope walker so that he falls from his rope, or to make performers juggle with imaginary objects. Music, public's reaction and comments, active tags which select the performers on the arena, contribute to the playful experience. All the video clips of the circus performers are organized in small loops, each serving the role of an elementary story fragment. The system will feature three performers and the public's reaction. The user interacts with the performer's video loops projected onto the table using hand gestures. In one case the participant will push the tightrope walker and make him loose balance. In another case he/she will spin the woman acrobat with a lateral gesture of the hand. Finally, the participant can interact with a contortionist and drive her into a full body knot. The demonstration currently uses infrared computer vision to track the user's hand and perform hand gesture recognition. What it is currently lacking is additional sensors for robust sensing, and a more sophisticated strategy to map hand gestures with the looped video segments. To provide an example of how sto(ry)chastics can help author the tabletop circus, I will focus on one of its featured digital performers, the tightrope walker, and will suggest a Bayesian network to handle sensor fusion and content selection. The proposed framework is only valid as an example, and further investigation will need to be carried out, as future work, to fully prove the expected results deriving from applying sto(ry)chastics to the tabletop circus.



Figures 184 and 185. Current prototype and setup for tabletop circus.

Here is a list of itemized video loops which are the elementary story fragments for the tightrope walker [figures 186,187]:

1. walk	5. kick
2. dance	6. bend
3. walk backwards	7. knee down
4. stand on one leg	8. spin around

The sensor modalities that we would like to fuse for this demonstration are: 1. Computer vision; 2. Capacitive sensing; 3. Doppler's radar to measure hand's velocity. The hand gestures/poses that are inputs to the system are:

1. point up	4. hand on table horizontally
2. point down	5. hand on table vertically
3. move sideways	6. grab with index & thumb, and drag

The meaning of the above actions in the context of this interactive multimedia experience are:

1. push sideways	7. poke
2. push down	8. pat/caresse
3. push up	9. do nothing or keep doing what you were doing before
4. halt	10. move away
5. help cross	
6. give kite	

Some of the above actions are clearly hostile, others are friendly, and therefore the user can be seen by the performer as friendly, or unfriendly, or neutral, which will influence the performer's response. The internal state of the tightrope walker which determine which content segment (video loop) is selected are:

1. happy	3. unstable (falling)
2. angry	4. neutral

From all the parameters listed above, it is obvious that authoring this interactive multimedia art piece can be quite a complex task. First of all there is only a subtle difference between some hand gestures which we need to distinguish from one another, such as for “poke”, “pat”, and “push”. Therefore the need for adding more sensors. The doppler radar can be added to make more accurate velocity measurements, which can be a key factor in disambiguating similar hand gestures. A capacitive sensor can also cooperate with vision in achieving a more robust position and hand shape information. The system needs further specification on how hand gestures, once correctly interpreted, affect the tightrope walker’s internal state, or how, in some cases, they directly determine his behavior. The hand actions also need to be interpreted in the context of the actual story being told. For example a “grab and drag” should not be confused with a “move sideways” gesture. Having a model for a friendly/unfriendly visitor can help interpret the user’s action without ambiguity or misinterpretation. A friendly visitor is more likely to be handing the kite to the tightrope walker whereas an unfriendly visitor will instead try to push the performer to make him fall.

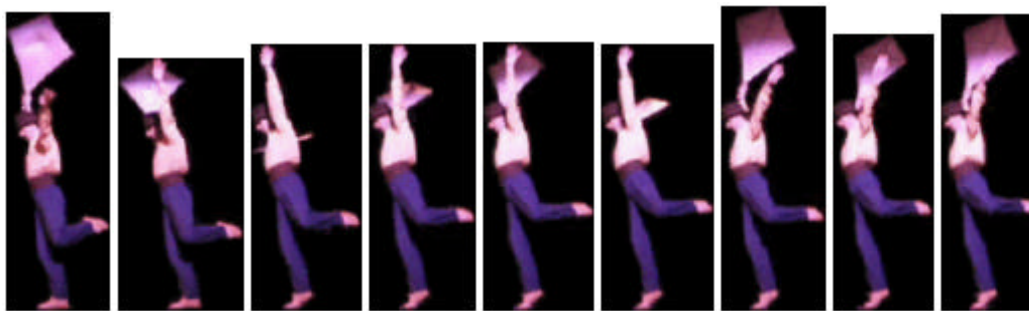
Given the high number of parameters that the author of the interactive experience need to take into account and control, using a traditional authoring system, such as any of the ones described in Chapter 2, can be a very difficult task. Even if the human author took care of choosing and encoding an appropriate heuristic for all possible cases, the resulting system would be inflexible to allow more video loops or more user actions to be added, because that would require specifying the input-output mapping from scratch. Nor such a system would allow the designer to easily try out various options or scenarios, as it is often needed when optimizing and fine tuning the application. The complicated web of relations and dependencies amongst all the system parameters is shown by figure 188. In addition to the interdependencies highlighted in the diagram, the author would have to account for the higher or lower degree of dependency amongst some of them, thereby multiplying the number of items to keep an accounting for, by the “resolution” of the dependency.

## Dance

---



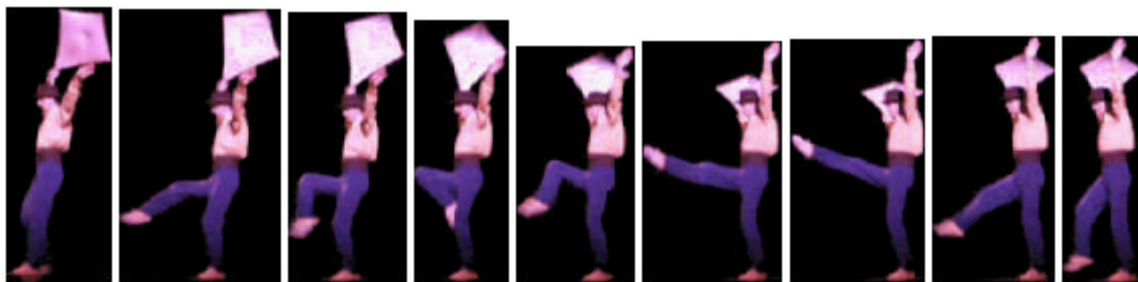
## Stand on one leg



## Knee down



## Kick



---

Figure 186. Representative frames for the tightrope walker in the tabletop circus.

## Bend



## Walk back



## Walk



## Turn

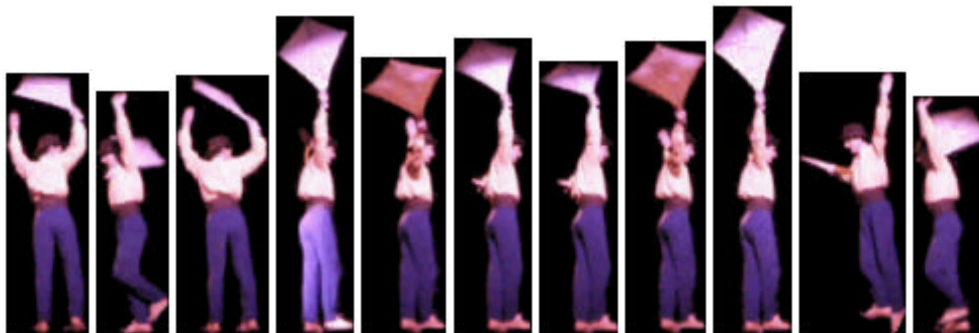


Figure 187. Representative frames for the tightrope walker in the tabletop circus.



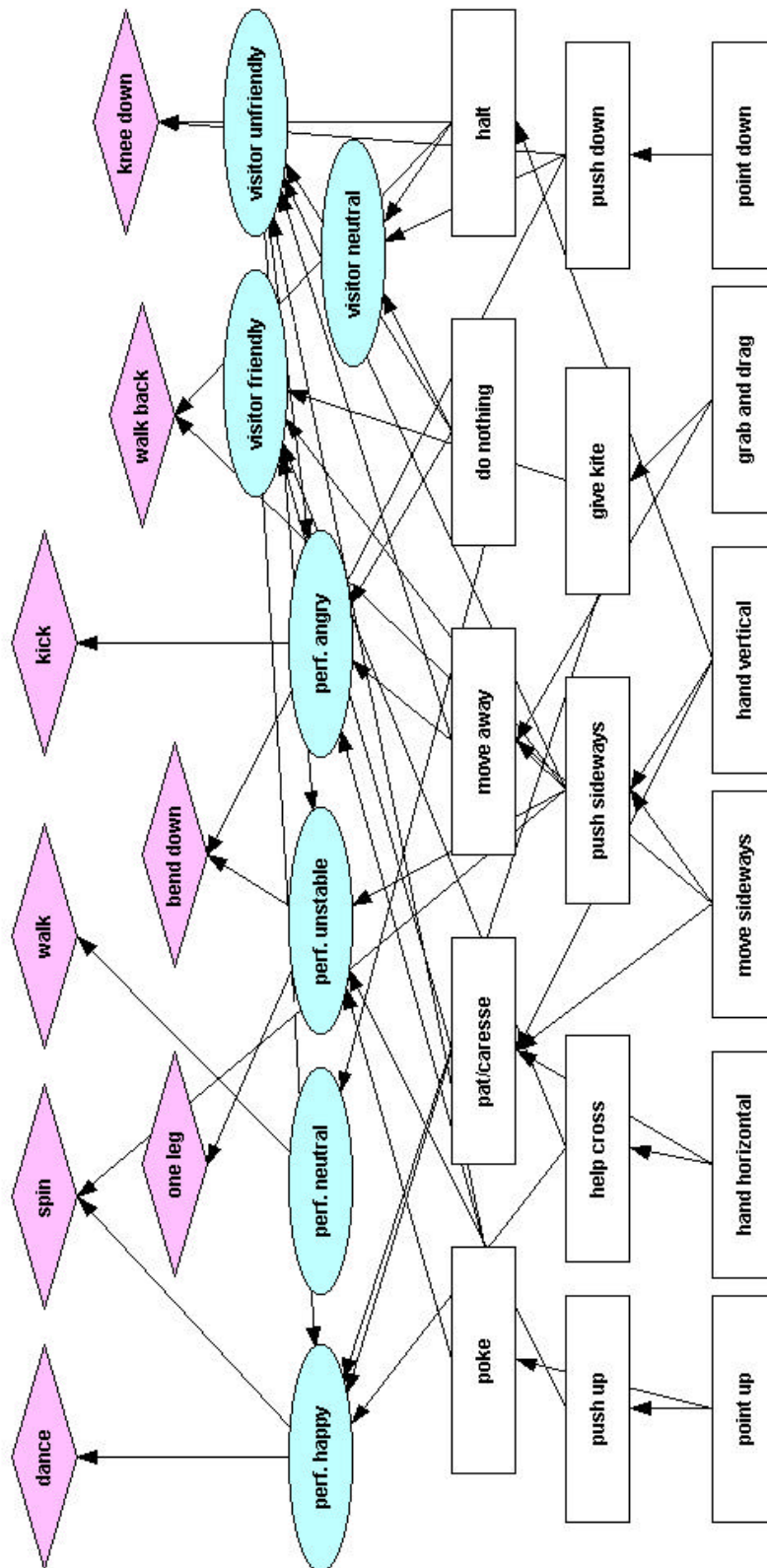


Figure 188. A scripted authoring for the tabletop circus would be very complicated



Another possibility is to simplify the interactive experience to fewer inputs and outputs to a one-to-one input-output responsive experience, which has been the approach of the majority multimedia authors so far.

A Bayesian network would allow the designer to handle the required authoring complexity and at the same time to ensure robustness in sensory data interpretation by way of sensor fusion. The steps that need to be undertaken to apply sto(ry)chastics to this problem are:

1. research a Bayesian network which correctly models the problem. Possibly try out various alternative models, and compare them, as shown in Chapter 4.
2. gather data of users interacting with the system
3. validate the model with the previous data. This can imply learning the parameters of the network proposed in step one, as shown in section 7.1., or going a step further, and learning the topology of the network from the data, if necessary.

Just as an example, or startpoint, an example of a possible Bayesian network used to model the tabletop circus as specified above is shown in figure 189. Figure 190 shows the possible internal states for some of the nodes. The root nodes for the sensors are not specified in the figure and are left blank. In some cases it may be best to have continuous nodes for the sensory information, whereas in others it is possible to classify the sensory information ahead into discrete categories, as for the museum wearable (see Section 7.1.). When the problem or data allows it, having all discrete nodes make the probability update and parameter/network learning easier and faster than having a hybrid network. Sometimes however this is not possible, nor desirable, and care has to be taken in making such modeling choices.



Figure 189. An example of a possible sto(ry)chastics authoring approach for the tabletop circus

