An Ambient Intelligence Platform for Physical Play

Ron Wakkary, Marek Hatala, Robb Lovell, Milena Droumeva
Simon Fraser University
2400 Central City
Surrey, British Columbia, Canada, V3T 2W1
604 268 7425
{rwakkary; mhatala; lovell; mvdroume}@sfu.ca

ABSTRACT

This paper describes an ambient intelligent prototype known as socio-ec(h)o. socio-ec(h)o explores the design and implementation of a system for sensing and display, user modeling, and interaction models based on a game structure. The game structure includes, word puzzles, levels, body states, goals and game skills. Body states are body movements and positions that players must discover in order to complete a level and in turn represent a learned game skill. The paper provides an overview of background concepts and related research. We describe the prototype and game structure, provide a technical description of the prototype and discuss technical issues related to sensing, reasoning and display. The paper contributes by providing a method for constructing group parameters from individual parameters with real-time motion capture data; and a model for mapping the trajectory of participant's actions in order to determine an *intensity* level used to manage the experience flow of the game and its representation in audio and visual display. We conclude with a discussion of known and outstanding technical issues, and future research.

Categories and Subject Descriptors

J.5 [**Arts and Humanities**]

General Terms

Documentation, Design, Experimentation, Human Factors

Keywords

Ambient intelligence, responsive environment, user model, physical play, puzzles, embodied, auditory display, motion capture, sound ecology

1. INTRODUCTION

This paper describes the research of an ambient intelligent system known as socio-ec(h)o. socio-ec(h)o explores the design and implementation of an ambient intelligent system for sensing and display, user modeling, and interaction models based on game structures. Ambient intelligence computing is the embedding of computer technologies and sensors in architectural environments that combined with artificial intelligence, respond to and *reason* about human actions and behaviours within the environment.

Ambient intelligent spaces lend themselves extremely well to physical and group play. In this paper we describe our interaction model and technical prototype. The overall research goal of this project is to understand to what degree physical play and game structures such as puzzles can support groups of participants as they learn to manipulate an ambient intelligent space. To date we have designed and implemented the prototype and interaction model. We have incorporated formative and summative feedback through a participatory design process and preliminary user testing.

The aim of our game is for a team of four players to progress through seven game levels. Each level is completed when the players achieve a certain combination of body movements and positions. At the beginning of each level, players are presented with a word puzzle as a clue in discovering the desired body states. The levels are represented by changes in the environment in light and audio. The levels are progressively more challenging in terms of body states and more complex in terms of the audio and visual ambient display. The physical environment consists of a circumscribed circular space (the area in which we can detect motion), immersive 8-channel audio, theatrical lighting, and two video projection surfaces, see figure 1.

The paper provides an overview of background concepts and related research. We then describe the game structure and prototype; include a technical review of the system, and a discussion of technical issues related to sensing, reasoning and display. We discuss our movement-based interaction and display in the context of aesthetic interaction. We describe how we utilized selective responses that were real-time, gradient, provided rewards, developed a composite model for reasoning on different groups of users, and how we customized a motion capture system for real-time bodily and positional sensing rather than gestures. We conclude with a discussion of known and outstanding issues and future work. Our contributions in this paper include a method for constructing group parameters from individual parameters with real-time motion capture data; and a model for mapping the trajectory of participant's actions in order to determine an *intensity* level used to manage the experience flow of the game; and design strategies for representing intensity via an audio and visual display.

Figure 1. Above are scenes from a user testing session

2. RELATED WORK

Key contextual issues in socio-ec(h)o stretch across many disciplines and research areas. It would seem that research in ambient intelligence in a social-cultural context is inherently interdisciplinary. The range of related topics include research in the areas of play and ambient intelligent spaces, motion capture systems, user modeling, auditory display, and literature linking play and learning.

Björk and his colleagues have observed progress toward fully ubiquitous computing games yet they identify the need to develop past end-user devices such as mobile phones, personal digital assistants and game consoles. Accordingly, we need to better understand how "computational services" augment games situated in real environments [4]. Interactive art projects such as works by F0am and Sponge have explored social and mixed reality environments incorporating gesture and wearable computing [15, 25].

Recent projects have investigated the play space of responsive environments and tangible computing utilizing sensors, audio, and visual displays. For example, Andersen [1], and Ferris and Bannon [10] engage children in exploratory play and emergent learning through sensor-augmented objects and audio display. Andersen's work reveals how theatrical settings provide an *emotional framework* that scaffolds the qualitative experience of the interaction. Ferris and Bannon's work make clear that a combination of simple feedback and control lead children to widely explore and discover a responsive environment.

In the *Nautilus* project, Strömberg and her colleagues employ bodily and spatial user interfaces as a way of allowing players to use their natural body movements and to interact with each other in a group game within a virtual game space [26]. Strömberg observed in physical and team games such as soccer or dodge ball that players coordinate their physical movements and rely heavily on communication to be successful. In their findings, participants reported that controlling a game through one's body movement

and position was "new and exhilarating." In addition, playing as a team in an interactive virtual space was found to be engaging, natural and fun.

In relation to the above research, socio-ec(h)o builds on the theatrical, simple and physical interaction models in order to develop a game structure approach that lies between exploratory play and a structured game for adults within an ambient intelligent environment. In addition, we extend the notion of a game structure to an interaction model for the environment rather than a virtual game space. We also build on the idea that action, play and learning are linked in such physically-based environments.

Technically, *Nautilus* employed a theatrical lighting approach similar to socio-ec(h)o. However, position tracking was done through the use of a sensor floor that tracked movement across x and y coordinates only. In socio-ec(h)o we utilized a motion capture system in order to support more complex actions and locations across three dimensions. Motion capture has primarily been used in contexts where 3D data can be captured for later analysis and re-processing. The technology is used with game construction and movie making to create realistic motions of animated characters. However, recently there is ongoing research into the real-time use of this technology in live performance and interactive arts [3, 18]. Particularly, motion capture was successfully used by part of the research team in a live dance performance titled "Immanence" where a dancer manipulated an animated figure in front of a live audience [14].

Another related topic to our research is user modeling in respect to ubiquitous computing and group models. Our work builds on adaptive and rules-based approaches to context-aware modeling [13, 22]. In addition, our approach has been influenced by previous research in group modeling [12, 17]. Specific to games research, we have utilized Richard Bartle's model of game types [2].

We have also benefited from the research in soundscape studies and acoustic communication through the works of Schafer and Truax [24, 27, 28], and in the area of cognition and psychoacoustics [5, 7, 19].

The approaches in this project, especially the further development of an ambient intelligent platform, and the use of user modeling and interaction models originated in earlier work by this research team on an ambient intelligent museum guide [11, 29].

Lastly, in respect to the links between action, play and learning there is a substantial amount of literature. Dewey argued for the *construction* of knowledge based on learning dependant on action [8]. Piaget, through his child development theory believes in the development of cognitive structures through action and spontaneous play [23]. According to Piaget, *constructivist* learning is rooted in experimentation, discovery and play among other factors. Papert extends Piaget's notions by investigating the knowledge-construction process that emerges from learners actually creating and designing physical objects [20]. Malone and Lepper consider games as intrinsic motivators for learning [16]. Subjective motivations like challenge, curiosity, control and fantasy may occur in any learning situation; other motivations like competition, cooperation and recognition are considered to be inter-subjective, relying on the presence of other players/learners.

3. GAME STRUCTURE AND PROTOTYPE

Below is a short scenario of participants in the socio-ec(h)o environment:

Madison, Corey, Elias and Trevor have just completed the first level of socio-ec(h)o. They discovered that each of them had to be low to the ground, still, practically on all fours. Once they had done that, the space became bathed in warm yellow light and filled with a wellspring sound of resonating cymbals. Minutes earlier, the space was very dim – almost pitch black until their eyes adjusted. A quiet soundscape of "electronic crickets" enveloped them. They discussed and tried out many possibilities to solving the word puzzle: "Opposites: Lo and behold." They had circled the space in opposite directions. They stood in pairs on opposing sides of the space. At Corey's urging, the four grouped together on the edge of the space and systematically sent a player at a time to the opposite side in order to gauge any change in the environment. Nothing changed. Madison, without communicating to anyone realized the obvious clue of "Lo" or "low". While Corey was in mid-sentence thinking-out-loud about the puzzle with Trevor, and trying to direct the group into new body positions, Madison lowered herself to a crouching position. The space immediately glowed red and became brighter. The audio changed into a rising chorus of cymbals – not loud but progressively more pronounced. Corey and Trevor stopped talking and looked around at the changing space. Madison, after a pause began to say "Get down! Get down!" Elias stooped down immediately and the space became even brighter. Corey and Trevor dropped down in unison and the space soon became bathed in a warm yellow light like daylight. The audio reverberated in the space. A loud cheer of recognition came from the group, "Aaaaahhh! We got it!" Corey asked everyone to get up. As soon as they were all standing, the space became pitched black again. They dropped down again and the space was full of light. They had learned how to "create daylight" in the space. They had completed level one.

Soon after, a new word puzzle was presented to them in a short video projected on two scrims hanging from the ceiling: "The opposite of another word for hello but never settles." The lights have become very dim now and the audio has a slightly more menacing quality to it. Level two will clearly be more challenging…

We formalized our game structure into a schema of levels, body states and goals, see table 1. As earlier described, the game has seven levels. The body states are the body movements and positions that players must discover in order to complete a level. Goals are the change in environment players are aiming to achieve. The goals are implicit and not explicitly stated for the players. Each level has a beginning quality of light and audio. As the players progress toward achieving the right body state, the environment incrementally shifts toward the goal state of the environment. For example, as depicted in the scenario, when Madison lowered herself, the environment gradually shifted toward the goal of creating day. As each of the other three players followed Madison, the environment responded to movements of each player.

Table 1: This table describes the socio-ec(h)o game schema.

Theme	Levels	Body State	Goal	New Game Skill
Discovery of light	1	"high-low"	create day	body position
Day for night	2	"moving low"	create night	movement/ duration
	3	"loosely moving"	create day	proximity
Rhizome	4	"dense center - scattered edge"	create spring	sequencing
	5	"this way slow – low to high"	create winter	sequencing/ duration
Biota	6	"two low moving – two high"	create summer	composition
	7	"ringing around the rosie"	create fall	composition & location

In addition, the schema includes new game skills and themes. We assigned each level a generic skill in relation to each body state and level. Despite the specific body state, the generic skill acquired at each level is required in order to discover the more complex body states at higher levels. For example, the specific body-state in level 1 is to crouch down low or sit down. The generic attributes learned are: being low to the ground and moving or not moving. In level 2, these generic skills are used to achieve the desired body-state of crawling or moving low on the floor.

Themes allowed us to design an implied progressive narrative based on natural evolution. Again, the specific themes and even the narrative are not known to the participants, rather they provide an underlying structure for body states, goal states and game skill acquisition. We intend for the progressive narrative to provide a sense of coherency across the levels, and to loosely map increased challenge to the reward of a more complex display. The content of our display systems, including light configurations and sound material also contributed in the higher levels toward creating and extending a narrative context.
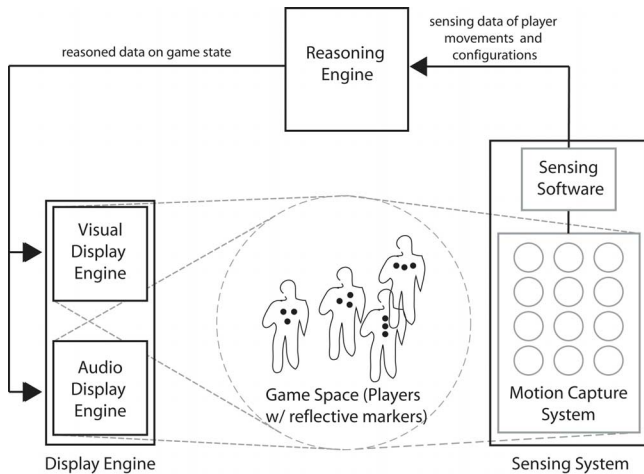
4. TECHNICAL PROTOTYPE

The technical system for socio-ec(h)o includes three key components, a sensing system, reasoning engine and display engine, see figure 2.

4.1 Sensing System

The sensing engine is comprised of a twelve-camera Vicon MX motion capture system (www.vicon.com) (see figure 3) and a custom program written in Max/MSP. Each participant is differentiated by unique configuration of reflective markers worn on their backs. Data is transmitted to the reasoning engine for high-level interpretation.
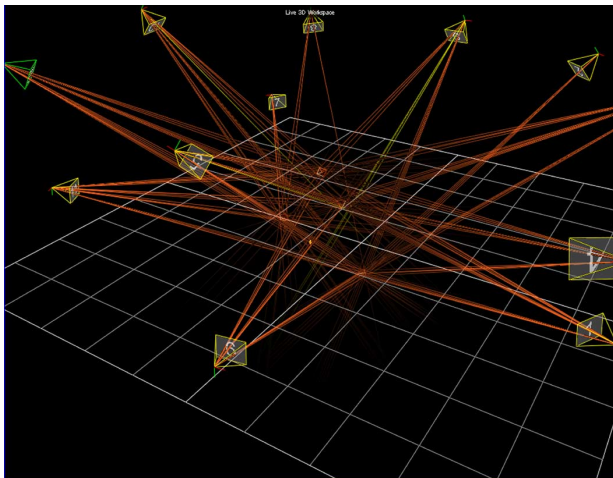
The motion capture system data was extracted via a proprietary protocol that Vicon uses to pass data packets between machines. Two Max/MSP objects were written that processed the information needed to allow the system to make decisions about game play.

Figure 2. This diagram depicts the system architecture. The architecture includes a sensing system, reasoning engine and display engine.

Information is extracted in two steps, first for individuals and then for group activities. For individuals, the system deduces from tag movement the following parameters: "low/high", "middle/outside", "fast/slow/still", "near someone/not near someone", "traveling/stationary", "direct/indirect motion, velocity", "location", "direction", "facing north-south/east-west/horizontal" and "visible/hidden". For each of these parameters a duration value was calculated, see table 2.

For groups, parameters are brought together into sets of activity. These are constructed based on each of the participant's individual parameters. For example, one set keeps track of who is low, while another keeps track of who is high. These activity sets are precursors for body states and help to determine the group parameters since body states look only to group behaviour. Because the associations are by parameters, one set for each parameter value is needed.

Figure 3. Graphical view of the Vicon motion capture system. Twelve cameras are used to reconstruct unique marker sets three dimensionally. These marker sets are used to identify and locate tags by the system. Here, four tags are shown with the camera rays that reconstruct their locations.

Table 2: This table details the parameters, threshold ranges, timing, and values utilized in our sensing system.

Parameter	Threshold	Timing	Values
Activity	n/a	n/a	Visible/Lost
Level	700-900 mm	n/a	High/Low
Speed	0 & 1.5 mm/sec	2 sec	Still/Slow/Fast
Space	140-170 mm	2 sec	Stationary/Travelling
Position	600-800 mm @ 0,0,0	n/a	Middle/Outside
Path	2-3 changes	2 sec	Direct/Indirect
Orientation	.5 radians	1 sec	N-S/E-W/Horizontal
Density	600 or 1250mm	1 sec	Loose/Dense
Duration	n/a	4 sec	Short/Long

Relational associations between people are another set of activities that involved determining what sets relationships have formed among the players. Since this involves measuring relationships between people instead of direct parameter measurements of individuals, the number of sets varies depending on how the associations form and un-form. With four players, this means that up to two associations could be active at any given time. One final set tracked who is not in a relational association with others in the group. *Density* is the only relational parameter tracked and it is based on the proximity of players to one another.
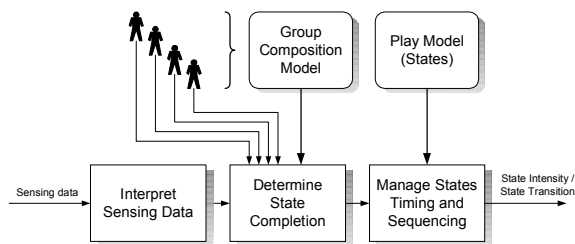
4.2 Reasoning Engine

The reasoning engine provides the intelligence for the system. It interprets the sensing data samples in real time, identifies the level of body state completion, and manages the narrative flow of the experience, see figure 4. The engine receives sensing data from the sensing system and interprets it in terms of high-level group behavior. For example the sensing system sends data on predefined parameters such as velocity and body positions. Based on these basic parameters the reasoning engine infers higher level parameters for the user group such as high-fast-moving group, middle-low-stationary group, etc.

The group parameters are further evaluated with respect to the individual user player types and the group composition model (group user model) that is dependent on the combination of user types as identified by Bartle's classifications [2]. As a result, each state completion is determined by its own function that depends on both individual and group characteristics. The function computes a single value we call state 'intensity' which is sent to the display engine.

Another role of the reasoning engine is to manage the flow in the game by sequencing of the states using the interaction model that defines the states and their sequencing at each game level. The single state levels simply require the group to complete the state in order to progress in the game. Multiple state levels require a group to complete a sequence of states within a certain time limit. The engine manages the timing of the state and level transitions.

The reasoning engine is rule-based and allows seamless modification and extension as the game evolves or expands. The reasoning engine feeds its output, state intensity and state transition to the display engine.

Figure 4. The processing chain for the reasoning engine. The individual user player types and a group composition are determining factors for state completion.

4.3 Display Engine

The display engine has two components, an audio and a lighting component. The audio display engine for socio-ec(h)o provides a sound ecology for each individual level of the system. It is custom software programmed in Max/MSP. We developed and structured the audio content on the principles of acoustic ecology and feedback-as-communication [27]. In addition, the audio display provides a gradient response to the participants, telling them how close they are to achieving their goal. The audio display system can alternate between stereo and multi-channel formats and localized and ubiquitous sound. The audio content follows the theme of evolution by utilizing sampled sound and several different sound processing techniques creating a shifting ambient soundscape that moves from simple, abstract sound to rich, environmental sound.

Lighting is manipulated with a DMX 512 controller via a Max/MSP patch. A small light grid and theatrical style lighting instruments and color scrollers are used. A lighting console was created to control multiple lights and color in concert through a cue list mechanism. Cues were written to simulate the various themes at each level.

Both the audio and the lighting systems take their cues from the reasoning engine, and respond to game aspects and configurations specified in the reasoning engine. Thus, the response of the display systems can potentially be used to provide feedback based on a variety of parameters such as how well participants are working together as a group.

4.4 Integration

Integration of the three components is achieved by lightweight communication protocol that is transferred over the User Datagram Protocol (UDP) communication channel. We consider uni-directional UDP communication appropriate for real-time applications.

5. TECHNICAL ISSUES

5.1 Sensing

In constructing a system to follow activities within the socio-ec(h)o space the main restriction on the sensing system is that it acquire information in a transparent way for participants. The system must also be able to uniquely identify individuals, track accurately in three-dimensions, and provide immediate response. These constraints led to the choice of a motion capture system because of the speed of response and ability to accurately and reliably track movement. While we have used general video sensing systems in the past [29], the challenge of performance efficiency, accurate individuation of multiple objects and high resolution three-dimensional tracking in a real-time environment proves too challenging despite advances in recent research [6].

However, the choice of the motion capture system restricted our movement space to an area that could be covered by twelve cameras – approximately 15 feet in diameter. In addition, several problems needed to be overcome. The Vicon motion capture system uses cameras to track reflective markers in the near-infrared spectrum (www.vicon.com). The system is a "passive" system, meaning that the markers do not send a signal that identifies the marker's identification. This is in contrast to an "active" system where markers can be tracked based on a signal it sends to the system either through a blinking rate or through an electromagnetic signal. Both systems are vulnerable to occlusion of markers by objects in the space that come between the marker and the cameras.

A passive system tracks markers in space and over time through calibrated models of where the markers are located in relationship to each other. Two approaches allow the markers to be identified and associated with the particular person to whom it is attached. First, the system tracks very quickly so that the movement from frame to frame is minimized. Second, the system has a model of what kinds of motion are possible. This forms a constraint model that the system can use to eliminate false positive matches of markers to people. The model contains information about how the markers are positioned in relationship to each other such as distance and joint styles.

Normally, motion capture systems are used to track the locations of limbs of one person as they move. However, with socio-ec(h)o we needed to track four or more people. This requirement is very difficult for a passive motion capture system to accomplish in even constrained environments. Some experimentation verified that the system easily confused movements of multiple people. It would often exchange arms of participants and even legs for arms, heads for hands etc. The system could track individual people for a short period of time if they stayed away from each other.

However, we decided not to track shapes of bodies moving in space but instead to track locations of where people moved. Identification of who is moving together with how they are moving is too difficult for a passive motion capture system to accomplish within the unconstrained socio-ec(h)o environment.

The motion capture system uses relationships of markers to identify individuals. A concept called a "rigid body" is used. Here, the system is told that certain markers will move in concert together, never changing their positions relative to each other (Another technique used to track markers is to specify how markers can move by the types of joints that are between the markers. For instance a hinge joint only allows movement back and forth through an angle).

A multiple marker identification system was created that allowed the system to track people in the space. We used four unique rigid body configurations of five markers each to identify and track participants as they moved. These marker configurations became identification tags, which are worn on the back. Five markers are used so that if any of the points became occluded the system might still be able to match its model to the data it is receiving. The more points that are used by the model the fewer false positives matches occur. Tags were created out of poster board Velcro and the markers supplied by Vicon, see figure 5.

Figure 5. Tags consist of five reflective markers in a configuration that is unique, independent of orientation.

The type of information extracted was qualitative in nature. We needed to have the system evaluate game play within human level perceptions of differences between things. It was decided that for group activities this would be information extracted in dualities such as high/low or very course categorizations such as fast, medium, and slow.

The tags were calibrated into the system prior to game play and remained reliably recognizable by the system. Overall, the system worked extremely well given the moderately constrained nature of the environment and the situation of game play. Information was extracted close to 30 frames per second but used by the reasoning engine at a rate of approximately 10 frames per second. It was found that this was adequate for game play.

5.2 Reasoning

The three major processing steps in the reasoning engine were described in section 4.2 above. The reasoning engine was implemented in Jess (http://herzberg.ca.sandia.gov/jess/). As the real time interpretation of large quantity of sensing data is required, careful memory management and activation/deactivation of the rules involved in the particular state are needed. The rule base for the engine is designed with this effectiveness in mind where only those group parameters that are needed for determining the completion level of the active state are computed. The engine also cleans all the history data in each processing step.

The rules for the system were hand-made and are specific for our set of sensed parameters and play states. There are 61 rules overall. The sample rule in Jess language that computes the group parameter for a "loosely-moving" body state is shown below where the rule preconditions are before '=>' symbol and actions follow:

```
(defrule group-parameter-high-stationary
  (or (goal-state ?l ? scattered-edge)
      (goal-state ?l ? two-low-moving-two-high-stopped))
  (current-time ?t)
  (group-parameter (users ?u1) (parameter space)
     (value stationary) (intensity ?p1) (time ?t) (duration ?d1))
  (group-parameter (users ?u2) (parameter altitude)
     (value high) (intesity ?p2) (time ?t) (duration ?d2))
=>
  (bind ?gr (group-intersect ?u1 ?p1 ?d1 ?u2 ?p2 ?d2))
  (bind ?users (explode$ (nth$ 1 ?gr)))
  (assert (group-parameter (users ?users)
          (number-of-users (length$ ?users))
          (parameter high-stationary) (duration (nth$ 3 ?gr))
          (time ?t) (intensity (nth$ 2 ?gr))))
  )
```

Theoretically, we utilized Bartle's concepts of collaborative play in Multi-User Dungeons (MUDs) and MUD Object Orienteds (MOOs) to help us formulate a group user model to support the reasoning within our system [2]. Bartle identified four types of MUD player styles: *achievers*, *explorers*, *socializers*, and *killers*. *Achievers* seek in-game success, *explorers* satisfy their environmental curiosity, *socializers* value human interaction, and *killers* exercise their will at the expense of other players.

The user and group models in socio-ec(h)o are static. The user player types are determined using Bartle's test [2] and entered into the system at the beginning of the play. A group composition model is established based on the individual player types. Both the individual player types and group model determine the dynamic of the system response to the group actions in the space. For example, for the group with a strong presence of 'explorers' the system response concentrates more on rewarding a required combination of body states that leads directly to the goal state (to 'curtail' the exploration). Alternatively, for a group with significant 'achiever' mix the system rewards each discovered partial body position that contributed towards the goal more strongly. For example, the rule computing the intensity level for group parameter 'middle', which requires the users to concentrate in the center, modifies the intensity value for each 'explorer' type user in the 'middle' with the following formula (4 is the maximum intensity):

$$intensity = intensity + (4\text{-}intensity)*0.2$$

Similar adjustments are applied for other user types and game states. As a result, the display response in socio-ec(h)o is adjusted with respect to the group composition. We are currently investigating our hypothesis that by considering the Bartle types participants have a better experience and more quickly become *skilled* interactors.

As mentioned earlier in section 4.2, the reasoning engine manages the narrative and experience flow of the interaction. The model for this includes mapping the trajectory of the body states to participant's actions in order to determine the *intensity* level, or proximity to the desired body state. The *intensity* level is measured from 0 to 4 with 4 representing the maximal intensity or state completion. As shown above, the *intensity function* is computed by applying several rules that modify the *intensity* level with respect to the state completion and group composition. Therefore, the *intensity function* is not computed by a single formula but is defined by heuristics that are applied in full response to the current state of the game. This provides for a powerful mechanism that enables us to evolve and expand our framework without the need to modify the game mechanism.

In addition, the ability to sustain *intensity* levels are also monitored, typically a 4 second duration is required to either complete the state or a state within a multi-state sequence (for example see level 4 in Table 1 in section 3.1). The overall shifts in intensities toward and away from the goal must be represented in a gradient effect yet be sufficiently real-time in order to best support actions in the environment. This overall model is utilized for sending data to the display engine in a managed flow, see Figure 6. In addition, the reasoning engine modulates the transitions from one level to another.
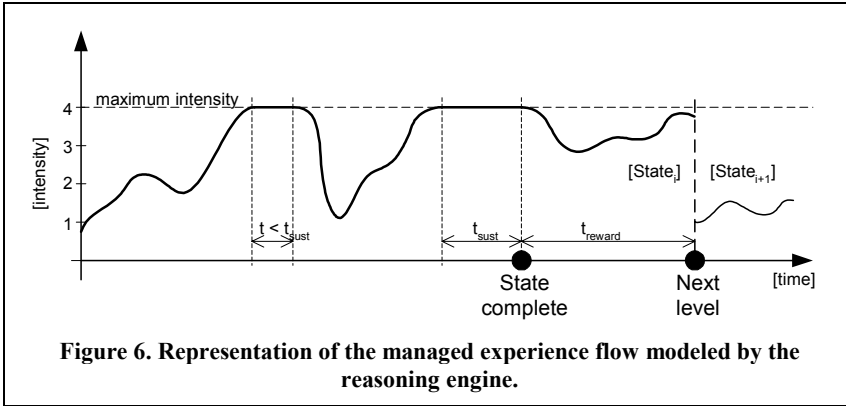
Figure 6. Representation of the managed experience flow modeled by the reasoning engine.

5.3 Display

5.3.1 Audio display

The audio display system provides an ambient, immersive auditory space that envelops participants in the play world and signals proximity to achieving the goal state of each game level. The system has three types of responses: an ambient, gradually changing soundscape that is distinct at each level; an intermediate reward sound to signal when all participants are working together towards their goal; and a final reward sound, which signals progressing to the next level. These responses work in conjunction with each other and complement the communication and aesthetic aspects of the users' experiences. In terms of Schafer's classification of sounds belonging to a local sound ecology, the three types of responses from the audio engine could be considered *keynote* sounds, *sound signals* and *soundmarks*, respectively [24].

The basic idea of our audio display system is to signal participants' success gradually, by *intensifying* the environment. Since the gradient response is key in this system, most thought had to be put in choosing the approach to representing intensity, as well as in the selection of the content both from aesthetic and cognitive perspectives.

The system uses several approaches to gradient response. The first is a simple cross-fader between 5 layers of sound which could be arbitrarily chosen to represent increased intensity of the same group of sounds (dripping water gradually changing into a fast river stream, over 5 steps). This type of soundscape design conforms with Truax's ideas of *variance and coherence* [27] – the balance of sameness and diversity, a core idea in perceptual design techniques. In other words, we use a group of sounds that are the same or very similar according to their basic characteristics of pitch, rhythm and timbre, and we represent success rate in the game by *intensifying* these basic sound characteristics.

The second step in representing intensity is realized through processing that is used to *colour* sound (slightly alter its core characteristics). We have used several different approaches in different game levels, based on previous research in auditory perception, and starting from "easier" to more subtle perceptual responses, see table 3. As research in classical and contemporary psychoacoustics suggest [19], amplitude change, followed by pitch change and tempo change are the most readily and easily perceived sound variance characteristics. Thus game levels 1 through 3 are based on these approaches of firstly *variance and coherence*, and secondly to *colour* sound. This allows us to start

with more contrasting change that eventually allows the listener to perceive more subtle changes. With tempo change we also benefit from Bregman's studies on auditory streaming which are made even more complex when using culturally significant sounds such as musical or environmental sound [5]. This is why intensity under this approach was related inversely – going from fast to slowing down with increased game success rate, to full, continuous sound at the completion of the goal state. This type of real-time smooth tempo change was realized by using a phaser.

Another approach to representing sound that was introduced as an evolution of previous levels of perception was using a low-pass bi-quad filter. This type of auditory perception is based in timbre differentiation, and while it entails a subtler, less precise representation, it proved quite useful in our system. The common recognition of this sound change was the feeling of "opening" and "closing" of a filter, a sliding from muffled sound to sharp, bright sound through attenuating different frequency bands in a given sound. Another useful technique, which seemed to help in better perceiving change, rather then representing change itself, was the use of 8-channel spatialization of sound layers. This was done by gradually moving the sound in a circle giving the users the impression that sound is "going away" or "getting closer" to their relative position in space. The reason why we thought this approach aided in perception was because the combination of localization and sound change would be greater than sound change alone, a notion, again suggested by some current studies in virtual audio [19].

Table 3. Audio display methods used at different game levels[1]

Level	Method	Description of effect	Perceptual aims
1	Bi-quad Filter	Change from muffled to bright sound	Perception of timbre
2	Phaser + Layer Fader	A gradual cross-fade of 5 sounds and tempo change	Tempo and timbre perception
3	Layer Fader + Pitch Shift	Cross-fade between 5 sounds (crackling sound)	Pitch and timbre perception
4	Layer Fader + Bi-quad Filter	Cross-fade 5 sounds and change from muffled to bright (fire sounds)	Perception of timbre and cultural semiotics
5	Phaser + Pitch Shift	Gradual pitch shift low to high, inverse tempo relationship (fast to slow)	Perception of pitch, tempo and associations
6	Layer Fader	A cross-fade of 5 sounds (calm forest – thunder)	Perception of timbre and associations

The final levels of the game, though represented simply by layering and cross-fading sound elements, had an added dimension to them, because as environmental sounds they held recognition and thus created narrative connotations for the game participants, helping them listen to and analyze the change in sound better.

---

[1] For testing, we only implemented six of the seven conceptualized levels depicted in table 1, section 3.

Our approach to reward sounds, although perceived as "signals" and thus analogous to direct feedback systems, were still a result of a "long-term" composite group action, rather than immediate individual reward. The reward sound signaling the passing of a level is perhaps the only sound mapped one-to-one to a particular event – the completion of goal state.

5.3.2 Visual Display

In regard to the visual display, two techniques are used to provide a gradient response to the participants. Initially, the only feedback provided was based upon intensity of the overall lighting in the room. As participants came closer to achieving the goal, the lighting moved toward a condition of light or dark that corresponded to the system's estimation of closeness to the goal. A direct mapping between intensity and goal closeness is used. While this is a useful technique for providing feedback, it did not provide much room for creating an ambient environment.

A second technique, allowed for environmental ambience and feedback to the participant by using a relative gradient between two states (two lighting states such as different colors and lighting levels for fall and winter). Transitions in the lighting were based on the value of the *intensity* function (see section 5.2). This is used to signal the lighting to transition toward the goal state over a fixed period of time (10 seconds). If the participant moves away from the goal, the lighting moves toward the other environmental state (the start state). At any point if there is no progress, or there is negative progress, the system moves toward the non-goal environmental state. When the goal state is achieved, then the lighting moves to the goal environmental state.

It was found that both techniques worked for providing feedback. Surprisingly, it was found that the second technique that used relative direction toward the goal provided the participants with a more satisfying experience. We believe that this occurred because the absolute state feedback is provided at a lower resolution of five states. Relative feedback was provided from between 5 to 32 differential gradations.

6. AESTHETIC INTERACTION

We aimed to emphasize the qualities of interaction that result in play that facilitates discovery and therefore explored the embodied and situated aspects of interaction or aesthetic interaction as expressed by Djajadiningrat [9] and Petersen [21].

In the domain of tangible user interfaces (TUI), Djajadiningrat argues for a "perceptual-motor-centered" approach [9]. He is less sympathetic toward the cognitive view of interaction in what he terms the "semantic approach" where objects communicate action through metaphor. We believe this approach equally applies beyond the TUI context to include gesture and movement-based interaction. The direct approach is governed by a "sensory richness and action-potential" to carry expression through interaction. We took guidance in his description of three factors as having a role in aesthetic interaction: the *interaction pattern* of timing, rhythm, and flow between the user and actions (and/or tangible objects); the *richness of motor actions* found in the potential space of actions and skill development; and *freedom of interaction* in which a myriad of interaction paths coexist. In addition, Petersen and her colleagues description of aesthetic interaction shares the embodied aspects described above as well as the sense of aesthetic potential that is realized through the action or engagement [21].

Our approach in the interaction model, visual display and audio display was to create action-potential interaction integrated with interpretive feedback space allowing for aesthetic potential in meaning and expression. For example, our approach to the sound content was significantly different from other audio-based immersive display environments in that it was neither entirely musical, nor was it entirely computer-generated synthesized abstract sound. Instead, the base consisted of sampled (field-recorded) sounds with a varying degree of abstraction and connotation, ranging from water and fire, to processed vocals and transients. We arrived at the final content elements through experimentation with composition of different iterations of sounds and sound processes and transitions, thus drawing from electro-acoustic and musical composition to create an engaging and an aesthetically rich experience for our participants.

The same degree of rich ambiguity and interpretive space was attempted with the word puzzles and embodied solutions. As we discussed earlier, the visual display experimented with qualitative experiences of intensities and gradients (see section 5.3.2).

7. PRELIMINARY USER TESTING

Our user testing to date is preliminary. It includes two three-hour sessions with eight participants, and an additional two-hour session with four other participants. All the participants were new to the game. The group included three females and nine males ranging in ages from twenty-one to fifty-nine. Two of the three groups had a gender mix. Each session began with a warm-up session to introduce the concept of puzzles solved through physical action and support through implicit responses. Participants were also played the range of sonic cues and rewards in order to attune their perceptual hearing to our sound ecologies – we found that prior listening is sufficient in creating recognition of reward sounds when they are heard again later in the game. Each team of four played two levels followed by questions and discussions. After all levels were achieved or a total of two hours of interaction (60 minutes in the shorter version), the game was stopped and a general open-ended interview and discussion took place. The first group participated in the environment for over two hours and completed four of the six operational levels. The second group completed all the levels in approximately ninety minutes. The third group completed three levels in sixty minutes. All groups were very engaged with the game and those who did not complete it wanted to continue.

Our post-play discussion was open-ended and focused on the overall experience and sense of game-play, collaboration and acquisition of game skills. We also pursued known issues or questions that were relevant to our stage of development. These included the perception of audio and visual thresholds, gradients and the role of abstraction or representation in the sound content; the appropriateness of our *intensity* function constructed through the reasoning engine (see section 5.2). Through observation and technical data we explored the applicability of our play types (see section 5.2), and known technical issues in relation to sensing and system performance. The results of these sessions serve the basis for our discussion in the following section.

8. DISCUSSION

Problems related to the sensing system did occur which in the end did not adversely affect the perceptions of the participants. While participants often occluded tags by walking out of the space

viewed by the cameras or through interactions where one person covered another's markers in each of the user testing sessions, these occlusions were brief and passed without disruption since the system utilized the last known position without consequence or the player's position was not relevant in progressing to a body state. Typical occlusions lasted less than two seconds but were observed for up to 30 seconds. Another issue with the system was the determination of divisions between categories, (threshold). Here the choice of a number required a calibration based on a subjective judgment of what is in one category or another (for instance what is slow and what is fast). It was found that careful choices of these parameters were adequate. However, in certain circumstances, slight adjustments were required to account for differences in abilities or body types.

Some of the known issues with the audio display include the mapping of game parameters to a number from 0 to 4 and its interpretation into sonic response. The almost complete lack of instant feedback as a departure from standard game design was a known issue and one we were exploring. Yet, the concept of *intensity* when using sound as representation is much less researched in the field of auditory perception than direct feedback-based response [15]. What we found in participatory design workshops and preliminary user testing was that this approach of gradual response was quite rewarding to the participants and encouraged their attentiveness towards the environment – both for light and sound display, interchangeably or complementary to each other. As well, our representation of intensity was largely successful as participants were able to identify their progress throughout the game based on the environmental cues.

As well, the audio content used in different levels had a varying degree of abstraction. In the course of our testing, it became clear that more abstract sound (such as a crackling of rocks) made it harder for participants to gauge the level of change in the gradient response, as compared to less abstract sound (such as fire or rain). Several of them suggested the feedback was not "crisp" enough, and our main extrapolation from this is that 1) they didn't perceive enough change and/or 2) they didn't recognize the sound and perceive its internal coherence or variance. The bi-quad filter turned out to be well received perceptually as an approach to gradual change in the sonic environment, while the phaser, depending of the content was subtler. What participants seemed to respond to most positively was a combination of environmental sound (whether intensifying fire or going deeper into the forest) and a multi-channel diffusion, rather than multiple stereo and abstract sounds. All participants commented on the immersion quality of the play space as a positive and rich experience.

As far as the intermediate and final reward sounds, the issues we were aware of had to do with masking, i.e. sound not heard over the ambience, as well as recognition of the sound, and appropriate content mapping between game state and reward. What we decided on was to use a random pitch variation of two soft abstract sounds (granulated tapping of glass). We also played the reward sounds to all participants before the game began to compensate for the fact that in our preparatory work it had been hard to distinguish the sound without a reference. This strategy seemed to be successful as people correctly identified the reward sound and used it effectively in their game play. On an aesthetic level it was hard to incorporate the seemingly intrusive, yet necessary, "reward" sound into the much more subtle and atmospheric soundscape environment. We considered the use of theme-appropriate rewards, that would be different for each level, but worried this will affect the rate of recognition of the reward, and ultimately – successful game-play.

Since we used a combination of approaches to representing intensity rather than a single approach per level, it is hard to generalize results about the perceptual effectiveness of individual approaches to sound processing. Yet we strongly believed that a combination of sound content and processing would work better than different approaches alone, and our preliminary tests supports that idea.

One of our intentions in this work was to explore the social interaction aspects of responsive environments. In this paper we have discussed the technical platform that we believe makes such an investigation possible. We plan targeted evaluation sessions focused on group collaboration, learning, communication and diverse approaches and goals within groups. In our discussion with testing participants three key observations arose: team collaboration represented by strategy-making and communication evolved considerably over a short period of time; communication was multi-modal including verbal, gestures and physical actions; leadership and decision-making typically rotated and all groups made an effort to share in the leading of the group to some degree, some groups more consciously than others.

9. CONCLUSION & FUTURE WORK
In this paper we have shown how our system builds on the theatrical, simple and physical interaction models. We have developed an ambient intelligence system that functions based on exploratory play with a conceptually structured interaction model. We feel the work contributes by providing a method for constructing group parameters from individual parameters with real-time motion capture data; and we provide a model for mapping the trajectory of participant's actions in order to determine an *intensity* level used to manage the experience flow of the game; and design strategies for representing intensity via an audio and visual display. We discussed related work in the areas of ambient intelligent game spaces, play and learning, motion capture based systems, user modeling and auditory display. We provided a description of our game structure and prototype and a detailed account of the technical system. We provided accounts of technical issues related to customization of the motion capture system, a rules-based and composite inference for reasoning on groups, and display issues of gradient responses, audio spatialization and real-time sonic generation and processing. We discussed our movement-based interaction and display in the context of aesthetic interaction. We detailed how the success of the experience relied on selective responses that were real-time, gradient, provided rewards and were unique to different group user models.

Future work includes a series of evaluations of the system to better understand the influence of the game structured interaction model, the supporting user model, and the display. In particular, we aim to understand how our approach enables a better experience and more *skilled* interactors within an ambient intelligent environment. Other work includes exploring the range and types of information that can be extracted from sensing. The work in this area involves mainly two avenues of approach: exploring new sensor paradigms and exploring types of

information that can be extracted such as more parameters or gesture based information.

10. ACKNOWLEDGMENTS

11. REFERENCES

[1] Andersen, K., "'ensemble': Playing with sensors and sound," in CHI '04 extended abstracts on Human factors in computing systems, (Vienna, Austria, 2004), ACM Press, 1239-1242.

[2] Bartle, R. "Players who suit MUDs," Journal of MUD Research, 1 (1). 1996.

[3] Benjamin, S., Mark, F., Hank, K., Luc, V., Lance, C. and Yu, H.-J., "Tele-sports and tele-dance: Full-body network interaction," in Proceedings of the ACM symposium on Virtual reality software and technology, (Osaka, Japan, 2003), ACM Press, 108-116.

[4] Bjork, S., Holopainen, J., Ljungstrand, P. and Akesson, K.-P. "Designing ubiquitous computing games – a report from a workshop exploring ubiquitous computing entertainment," Personal and Ubiquitous Computing, 6. 2002, 443–458.

[5] Bregman, A.S. Auditory scene analysis: The perceptual organization of sound. MIT Press, Cambridge, Massachusetts, 1990.

[6] Connell, J., Senior, A.W., Hampapur, A., Tian, Y.-L., Brown, L. and Pankanti, S., "Detection and tracking in the ibm peoplevision system," in Multimedia and Expo, 2004. ICME '04. 2004 IEEE International Conference on, (Taipei, Taiwan, 2004), 1403-1406.

[7] Cook, P. Music, cognition, and computerized sound: An introduction to psychoacoustics. MIT Press, Cambridge, Massachusetts, 1999.

[8] Dewey, J. Democracy and education: An introduction to the philosophy of education. The Macmillan Co., New York, 1938.

[9] Djajadiningrat, T., Wensveen, S., Frens, J. and Overbeeke, K. "Tangible products: Redressing the balance between appearance and action," Personal and Ubiquitous Computing, 8 (5). 2004, 294 - 309.

[10] Ferris, K. and Bannon, L., ""A load of ould boxology!"" in Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques, (London, England, 2002), ACM Press, 41-49.

[11] Hatala, M. and Wakkary, R. "Ontology-based user modeling in an augmented audio reality system for museums," User Modeling and User-Adapted Interaction. 2005, In Press.

[12] Ilenia, G., Fabio, P., Massimo, Z. and Dina, G.-B., "Dimensions of adaptivity in mobile systems: Personality and people's attitudes," in Proceedings of the 10th international conference on Intelligent user interfaces, (San Diego, California, USA, 2005), ACM Press, 223-230.

[13] Jameson, A. "Modelling both the context and the user," Personal Ubiquitous Computing, 5. 2001, pp. 29-33.

[14] Kozel, S., Lovell, R. and Griffiths, J. Immanence, Studio Hydro Quebec, Montreal Canada, 2005.

[15] Kuzmanovic, M. and Sha, X.W., "Weedy sociality | distributed wilderness," in Directions and Implications of Advanced Computing, (Seattle, Washington, U.S.A., 2002). http://f0.am/publications/2002_diac/index.html

[16] Malone, T.W. and Lepper, M.R. "Making learning fun: A taxonomy of intrinsic motivations for learning," in Snow, R. and Farr, M. eds. Aptitude, learning, and instruction: Cognitive and affective process analyses, Lawrence Erlbaum, Hillsdale, N.J., 1987.

[17] Masthoff, J. "Modeling a group of television viewers," User Modeling and User-Adapted Interaction, 14. 2004, 37-85.

[18] Meador, W.S., Timothy, J.R., Kevin, O.N., Eric, K. and Carol, C. "Mixing dance realities: Collaborative development of live-motion capture in a performing arts environment," Computer Entertainment, 2 (2). 2004, 12-12.

[19] Neuhoff, J.G. Ecological psychoacoustics. Elsevier Academic Press, Boston, 2004.

[20] Papert, S. Mindstorms: Children, computers, and powerful ideas. Basic Books, New York, 1980.

[21] Petersen, M.G., Iversen, O.S., Krogh, P.G. and Ludvigsen, M., "Aesthetic interaction: A pragmatist's aesthetics of interactive systems," in Proceedings of the conference on Designing interactive systems: processes, practices, methods, and techniques, (Cambridge, Massachusetts, U.S.A., 2004), 269-276.

[22] Petrelli, D., Not, E., Zancanaro, M., Strapparava, C. and Stock, O. "Modelling and adapting context," Personal Ubiquitous Computing, 5 (20). 2001, pp. 20-24.

[23] Piaget, J. and Piaget, J. To understand is to invent: The future of education. Grossman Publishers, New York, 1973.

[24] Schafer, R.M. The new soundscape: A handbook for the modern music teacher. BMI Canada, Don Mills, Ont., 1969.

[25] Sha, X.W., Y. Serita, J. Coffin, S. Dow, G. Iachello, V. Fiano, J. Berzowska, Y. Caravia, D. Nain, W. Reitberger and Fistre, J., "Demonstrations of expressive softwear and ambient media," in UbiComp 2003 Adjunct Proceedings, (Seattle, Washington, U.S.A., 2003), 131 - 136.

[26] Strömberg, H., Väätänen, A. and Räty, V.-P. A group game played in interactive virtual space: Design and evaluation. ACM Press, London, England, 2002.

[27] Truax, B. Acoustic communication. Ablex Publishing, London, 2001.

[28] Truax, B. Handbook for acoustic ecology. Cambridge Street Publishing, London, 1999.

[29] Wakkary, R., Hatala, M. and Newby, K. "Ec(h)o: Ecologies for designing playful interaction," in Frascara, J. ed. Designing effective communications: Creating contexts for clarity and meaning, Allworth Press, New York, NY, 2006, In Press.