AmbiLearn:
Ambient Intelligent Multimodal Learning Environment for Children

Jennifer Hyndman, B.Sc. (Hons.)
Supervisors: Dr Tom Lunney, Prof. Paul Mc Kevitt

Confirmation Report
June, 2009

Presented as a requirement for Ph.D. in
School of Computing and Intelligent Systems
Faculty of Computing and Engineering
University of Ulster, Magee
Email: Hyndman-j2@email.ulster.ac.uk

Abstract

Mobile technologies are demonstrating significant potential to aid and assist us in our everyday lives. Mobile phones and handheld computers enable us to communicate and gain access to information anywhere and at any time, while gaming devices are proving to be successful and a popular entertainment medium especially amongst children. This report presents research investigating the use of multimodal interfaces on mobile handheld devices, but more specifically, the use of a mobile multimodal learning environment for children is addressed. The primary objective of the research is the design and development of AmbiLearn, a multimodal learning environment for children based on a mobile handheld device. Using speech and 2D gesture as input and providing speech, graphics and an embodied agent as output, the development of AmbiLearn will address the interface design and usability of a multimodal system for children. Addressing the role of serious games in education, an application of AmbiLearn called TreasureLearn will be developed to investigate whether an educational game integrated with elements of a virtual learning environment based on a handheld gaming device can have a positive impact on children's education at primary level. Research into multimodal interfaces with a focus on handheld devices is provided, including modalities such as spoken dialogue systems and the use of embodied agents. Learning environments are discussed highlighting the use of games in education, virtual and augmented reality and the role of mobile handheld devices in this context. The project proposal and AmbiLearn are discussed via an architecture diagram and TreasureLearn application activity diagram, to give an overview of the structure of the proposed system. A comparison of features of existing similar systems with those of AmbiLearn is provided and a research plan is outlined. Potential tools are reviewed which may aid in the development of AmbiLearn and the conclusion provides information about the future direction of the research. This investigation aims to contribute knowledge to the fields of mobile HCI and more specifically child computer interaction, mobile multimodal interfaces and educational learning environments.

Table of Contents

1 Introduction

Ambient intelligence involves the convergence of several computing areas such as ubiquitous or pervasive computing, intelligent systems and context awareness with the added appreciation of social interactions of objects in the environment (Shadbolt, 2003). A key technology emerging from this field is intelligent user interfaces, particularly natural user interfaces. Intelligent user interfaces are multifaceted in purpose and nature, and include capabilities for multimedia input analysis, multimedia presentation generation, and the use of user, discourse and task models to personalize and enhance interaction (Maybury, 1998). The advancement of spoken dialogue systems (Mc Tear, 2004; Delgado et al. 2005; Wahlster, 2006; Bernsen et al. 1998) and multimodal interfaces (Maybury, 1993; Mc Kevitt, 1995/96; Oviatt, 2007) offer new ways of interaction replacing the traditional keyboard and mouse. Multimodality has real potential to improve interaction between computers and people. Research in the fields of pervasive computing and ambient intelligence has resulted in multimodal interfaces that enable communication between technology and people to be more natural and efficient. Pervasive technologies, such as the Personal Digital Assistant (PDA), smart phones and handheld gaming devices are prime examples of where multimodal interfaces are of benefit. However these devices are rarely used by children for educational use. With many organisations 'harnessing technology' for educational and entertaining purposes, there is a growing field of edutainment software for children. A vision of ambient learning, allowing a person to have access to information anytime, anywhere allows for a richer, flexible learning environment. The combined use of multimodal edutainment system based on a handheld device has the potential to provide a rich experience for collaborative learning with tasks not restricted to a physical location on a traditional desktop PC (Hyndman et al. 2009).

1.1 Research objectives

In recent years much research is investigating the potential of technology for use in education and terms such as personalized learning, intelligent tutoring and m-learning have brought significant advances within education institutes. However, the use of these methods may not have significantly impacted on educators at the primary level. This research thus aims to investigate whether an educational game integrated with a virtual learning environment based on a handheld mobile gaming device can have a positive impact on children's education at a primary level.

The objectives of this research are to:

- Investigate methods of merging multiple modalities on a mobile handheld device (PDA/Nintendo DS).
- Address the interface design and usability of a multimodal system for children.
- Consider the role of serious games in children's education.
- Demonstrate and test the educational potential of multimodal, mobile games (TreasureLearn).

1.2 Report outline

This report outlines a research project involving the design of an ambient intelligent multimodal learning environment for children, accessible via a mobile handheld device. Research will focus on child HCI and investigate the use of multiple modalities as a means of enhancing interaction between children and technology. Attention will focus on the use of multimodal communication on a mobile device to support collaborative learning with children. A multimodal system, called AmbiLearn, will be developed which combines speech and pen as inputs and graphics, embodied agents and speech as outputs. As a testbed application, TreasureLearn, an educational game will be developed, which integrates appropriate modalities to facilitate testing and evaluation of the system. The remainder of the report is structured as follows: Section 2 provides a review of research relating to key concepts within the fields of Multimodal Interfaces and Learning Environments. Section 3 gives an overview of the proposed project including the motivation for the work, the AmbiLearn architecture and the TreasureLearn application. AmbiLearn is compared to related systems and a project research plan is given. Section 4 documents prospective tools and software which may be used in the development of AmbiLearn. Section 5 concludes the report and outlines the future direction of the project.

2 Literature review

This chapter provides a review of research relating to key concepts within the field of multimodal interfaces, learning environments and serious games, all of which provide context for the development of AmbiLearn.

2.1 Multimodal interfaces

Multimodal interfaces are computer interfaces which enable more than one mode of interaction. Advances in hardware have enabled opportunities whereby interfaces are no longer restricted to the traditional keyboard and mouse input. Eye tracker technologies allow a users eye gaze to be tracked allowing the eye to focus on specific points such as text fields or on screen keyboard characters (Hutchinson et al. 1989; Charlier et al. 1997). This technology also provides valuable data used for evaluation purposes where the user data can identify characteristics such as, which features held the users attention, for how long was the gaze in a particular position, how the users react to stimuli and whether a natural interaction occurred (Witkowski et al. 2001; Santamaria, 2003; Prendinger et al. 2005). Touch screen displays, much like a traditional mouse, allows a user to directly manipulate or navigate the interface using a stylus. Using a stylus also allows easier input of drawn symbols, signs, arrows, lines and handwritten words to be recognised. Microphones are used to provide speech input and cameras for visual input. The visual input may consist of pointing to particular areas on a map or specialised surface demonstrated by Brøndsted et al. (2001) in *The IntelliMedia Workbench*, or 3D gestures such as those use in *SmartKom* (Wahlster et al. 2001). Systems such as Bolt's *Media Room* (Bolt, 1980) and work by Koons et al. (1993) outline multimodal interaction, with different modes of interaction. The *Media Room* contained a large projected screen and a chair with a touch sensitive joystick on each arm. Combined use of speech and gesture enables the user to say “put that there” whilst pointing to the ‘object' and ‘position' at the time of the utterances ‘that' and ‘there'. Koons et al. (1993) describe a prototype which accepts simultaneous speech, gesture and eye movement as input. Using a microphone, an eye tracker and full hand sensing hardware the user can manipulate an on-screen map using a free mixture of the modalities.

2.1.1 Speech and pen input

In her paper Ten Myths of Multimodal Interaction, Oviatt (1999) discusses that a user will not necessarily interact in a multimodal manner all of the time, but use each available modality when it is easier to carry out a certain task. With this in mind it is easy to identify why Cohen (1992) suggested that a major principle within the development of multimodal interfaces is to use the strengths of one modality to overcome the weaknesses of another. Table 2.1 outlines this principle, where it is evident that the strengths of using natural language, overcomes the weaknesses provided by using direct manipulation and vice-versa. These complimentary input methods are arguably the most common and much research has focussed on the combined use of speech and pen as input (Oviatt et al. 2000; Dusan et al. 2003; Buisine et al. 2005).

2.1.2 Multimodal mobile interfaces

The development of speech and pen as complimentary interface modalities is supported by their growing usage within mobile applications. A pervasive device needs an interface where the user is not restricted to a physical location as their hands or eyes may be occupied doing other tasks. Oviatt et al. (2000) provide an extensive review of speech and pen-based gesture systems and suggest that these complimentary methods amongst others, support improved efficiency, satisfy higher levels of user preference and accommodate a wider range of users, tasks and environmental situations. Many projects have focussed specifically on multimodality in mobile devices including *The Mona Project* (Anegg et al. 2004), *SmartKom Mobile* (Malaka et al. 2004) and *Quickset* (Cohen et al. 1997). The Mona Project (Mobile multimodal Next generation Applications) focuses on multimodal interfaces on mobile devices. The team have also developed the MONA presentation server supporting the modalities of speech, text, graphics and non-speech audio (Anegg et al. 2004). The server provides services independent of application, device and modality which are demonstrated by the *MONA@play* application (Baillie et al. 2005) and *MONA@work* (Niklfield et al. 2005). SmartKom Mobile runs on an IPAQ mobile device offering services such as route planning and interactive navigation. This system offers multimodal input through means of speech and 2D gesture using a stylus and a simplified version of Smartakus (SmartKoms interface agent) provides speech, gestures and facial expressions as output (Malaka et al. 2004). Quickset provides pen and voice input from a handheld device to distributed applications such as military simulations and medical information retrieval (Cohen et al. 1997).

	Direct Manipulation	*Natural Language*
Strengths	1. Intuitive 2. Consistent look and feel 3. Options apparent 4. Fail safe 5. Feedback 6. Point, Act 7. "direct engagement" with semantic object 8. Acting in "here and now"	1. Intuitive 2. Description, including; a. Quantification b. Negation c. Temporal information 3. Context 4. Anaphora (e.g., pronouns) 5. Delayed action possible
Weaknesses	1. Description, including; d. Quantification e. Negation f. Temporal information 2. Anaphora 3. Operations on large sets of objects 4. Delayed actions difficult	1. Coverage is opaque 2. "overkill" for short or frequent queries 3. Difficulty of establishing and navigating context 4. Anaphora is problematic 5. Error prone 6. Ambiguous

Table 2.1 Direct Manipulation and Natural Language as complimentary interface modalities (Cohen, 1992)

2.1.3 Children and multimodal interfaces

The use of multiple modalities is ideal within multi-user collaborative environments. A study on children's collaborative interactions (Scott et al. 2003) suggests that sharing a physical display with multiple input devices may improve collaboration due to a heightened awareness of the other user's actions and intentions. Multimodal interfaces such as *KidsRoom* (Bobick et al. 1998a, b) and *TICLE* (Tangible Interfaces for Collaborative Learning Environments) (Scarlatos, 2002) are intended for multi-user participation supporting collaboration. KidsRoom developed at the MIT Media Lab, is a fully automated and interactive narrative play space for children (Bobick et al. 2000). The environment uses computer vision and speech recognition to identify users actions and based upon these actions provides output in the form of images, video, light, sound, music and speech. The TICLE system 'watches' as children work together on a tangram puzzle (Scarlatos, 2002). TICLE uses computer vision to track the individual pieces of the puzzle. Based on the 'state' the puzzle is in, TICLE provides context aware feedback through speech and a graphical display.

2.2 Spoken dialogue systems

Spoken dialogue systems (SDS) provide a means for an interface to understand and provide spoken language. The main components of a SDS are outlined by Mc Tear (2004) as: Speech recognition, language understanding, dialogue management, language generation and text-to-speech synthesis.

2.2.1 Speech recognition / automatic speech recognition

Speech recognition is the process by which an acoustic signal, i.e. a spoken utterance, is translated into a sequence of words. Figure 2.1 outlines the process of automatic speech recognition (ASR) as described by Mc Tear (2004). The process basically involves extracting unknown features from an acoustic signal and matching these with a predefined set of patterns which have been established from training data. The best match between the extracted features and the predefined models is then presented as the most likely recognition sequence. The process of pattern matching involves using statistical models which are computationally expensive, i.e. they may require a large amount of memory and processing power.

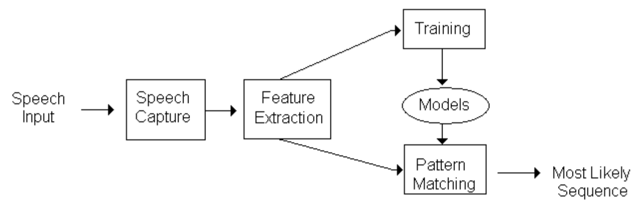

Figure 2.1 Automatic Speech Recognition Process (Mc Tear, 2004)

Due to the amount of processing and memory needed for ASR, limitations occur when using ASR on mobile devices. Zaykovskiy (2006) provides a survey of techniques for ASR on mobile devices. Three principle system architectures are highlighted which include: Embedded, Network and Distributed. With an embedded speech recognition system, the entire ASR process occurs within the mobile device. However, this has disadvantages due to the limitations of mobile devices regarding processing power and memory requirements. These embedded ASR systems are suitable for Personal Digital Assistants (PDAs) and mainly support small vocabulary sizes which are ideal when the utterances require single words or low level intentions. Network speech recognition (NSR) captures speech input on a client with limited computational power and performs complete recognition process on a server. Distributed speech recognition (DSR) systems process the speech in two parts. The feature extraction occurs on the client side (PDA) whilst the ASR search, which is the computationally expensive part, resides on the server.

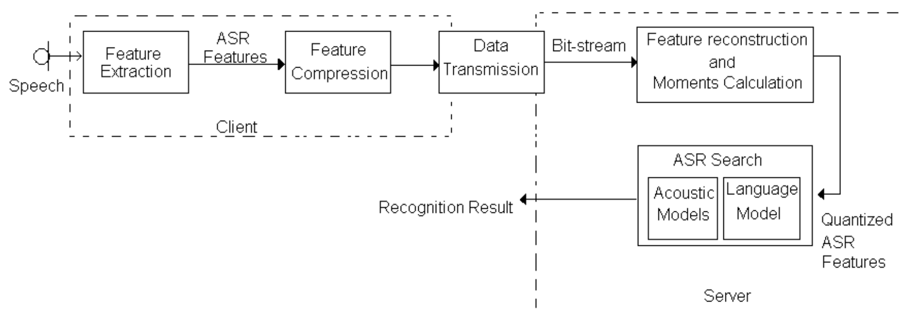

Figure 2.2 Distributed speech recognition (Zaykovskiy, 2006)

2.3 Embodied agents

The face provides an important means of providing non-verbal communication, such as raising an eyebrow, smiling and frowning. Eye movements show a person's interest or lack of it and the face transmits emotions and offers cues for natural turn taking protocols. Embodied agents are virtual representations of a person or character which provide and recognise non-verbal communication modalities. There are many potential benefits of having an embodied agent in multimodal systems which include: grabbing the user's attention, social interaction, naturalness and non-verbal feedback (Ortiz et al. 2007). Pedagogical Agents are embodied agents designed specifically for the purpose of teaching. The application of agents in educational software has many benefits as they can collaborate with the student. The Persona Effect (Lester et al. 1997) was an empirical study which outlined that well crafted lifelike agents have an exceptionally positive impact on students. Students perceived the agents as being helpful, credible and entertaining. Lester et al. (1997, p. 359) states, "the captivating presence of the agents can motivate students to interact more frequently with agent-based educational software. This in turn has the potential to produce significant cumulative increases in the quality of a child's education over periods of months and years."

2.3.1 Gandalf

An early example of an embodied agent is Gandalf, represented as a 2 dimensional face, hand and a voice (Thórisson, 1996). The technology involved allows Gandalf to perceive a users intention through speech using a microphone, body language and manual gesture through the use of a body suit and gaze through the use of a head mounted eye tracker. Gandalf, applied to information on the solar system, is capable of real-time multimodal dialogue responding to the user's intentions through speech, gaze, facial expressions and gestures as well as using a projected screen to display required information. Capabilities of Gandalf involve basic communication skills such as; back channel feedback, attentional cues and using communicative facial expressions and fillers (Thórisson, 1997). Figure 2.3 illustrates a user interacting with Gandalf (Thórisson, 2005).

Figure 2.3 Interaction with Gandalf (Thórisson, 2005)

2.3.2 Rea - "Real Estate Agent"

Rea is a full-sized agent capable of speech with intonation, facial gestures, body gestures and posture. As a real estate salesperson, Rea is aware of the functions of verbal and non-verbal conversational behaviours thus improving interaction by helping the user remain aware of the state of the conversation (Cassell et al. 1999). Rea senses the user's intentions using a microphone worn by the user and two fixed cameras which track the users head and hand movements.

2.3.3 Steve – "Soar Training Expert for Virtual Environments"

Steve is a pedagogical agent used for procedural training in virtual environments (Johnson et al. 1997). With the ability to give instructions and demonstrate activities Steve maintains an episodic memory of situations providing means to explain his actions when asked. Steve continuously monitors students' actions in the virtual world and so can offer context aware assistance when needed. Interacting in the virtual world involves the user wearing a head mounted display which gives the user a 3-dimentional view of their surroundings. A microphone is used to talk with Steve and data gloves or positioned sensors on the hand, allow the user to interact with the virtual machinery (Rickel et al. 1999). Figure 2.4 illustrates Steve highlighting a certain feature to the user (USC, 2000).

2.3.4 Hans Christian Anderson (HCA)

Within the NICE (Natural Interactive Communication for Edutainment) project (Corradini et al. 2005), an embodied conversational character namely Hans Christian Anderson (HCA), was used to interact with children. The children could ask HCA questions about his life, historical period and his fairytale characters through natural language. Using spoken language and 2D gestures, the user can follow HCA around a fairy tale land and gain context aware information. Figure 2.5 highlights the HCA character that resides within the NICE project (Martin et al. 2006).

2.4 Human computer interaction

The success of any system or new technology is largely dependent on whether the user accepts it. For this reason HCI still plays a major role in all technology related research. HCI is related to the study of interaction between human and computers and as new technologies emerge, new interaction styles and modalities become available. Child-Computer Interaction (CCI) (ChiCI group, 2007) is a sub field of HCI and its attention is focused on the interaction between children and technology. Research is this field is relatively limited compared to that of mainstream HCI, however just as important when the user group targeted is children. As a growing field CCI researchers share a common belief that the usability of technology for children is worth special consideration (ChiCI group, 2007). Bruckman et al. (2007)

identified several characteristics which differ between children and adults and therefore need to be considered when developing systems for children. These include: dexterity, speech, interaction style, reading, and background knowledge. Strommen et al. (1996) studied the use of different pointing devices with children. Comparing a joystick, mouse and trackball, Strommen concludes that adult standards for an optimal interface are not appropriate when children are the intended users.

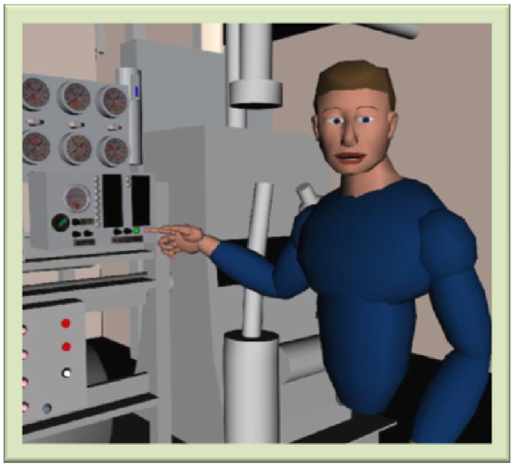

Figure 2.4 Interaction with Steve (USC, 2000)

Figure 2.5 The NICE project's HCA character (Martin et al. 2006)

2.5 Learning environments

"I never try to teach my students anything, I only try to create an environment in which they can learn." - Albert Einstein

Learning environments are traditionally perceived as an instructor based process within classrooms. However technology has provided opportunities where learning migrates beyond the classroom. The advent of the World Wide Web and a vision of ambient learning, allowing a person to have access to information any time, anywhere, allows for a richer, flexible method of providing the learning environment.

2.5.1 Technology enhanced learning

The term technology enhanced learning is used to cover all methods of using technology as a foundation for learning such as: m-learning, e-learning and web-based learning. Although these are different, they provide a common method for using different technologies to provide the learning environment. O'Malley et al. (2003, p. 6) defined m-learning as "Any sort of learning that happens when the learner is not at a fixed, predetermined location, or learning that happens when the learner takes advantage of the learning opportunities offered by mobile technologies." Similar to this definition e-learning is electronic learning where the learning is based upon using some form of technology but not limited to mobile technologies. The terms here are broad and vary depending on the technology being used, however a common theme amongst each is the method of providing the pedagogical content which makes it effective. Moreno and Mayer (2007, p. 310) provide an investigation of interactive multimodal learning environments. They suggest that "the most effective learning environments are those that combine verbal and non-verbal representations of the knowledge using mixed-modality presentations". Based upon principles derived from cognitive theories of learning, Moreno and Mayer (2007) summarise principles that apply to designing interactive multimodal learning environments outlined in Table 2.2.

2.5.2 Games in education

A growing field of interest within education is the use of games as a method of learning. A recent EU report suggests that 'Video games are good for children'. Summarised by Booth (2009), the report outlines that games promote the development of skills such as strategic thinking, creativity, cooperation and innovative thinking. Providing a fun, playful approach games can facilitate learning in three ways: learning as a result of tasks stimulated by the content of the games, knowledge developed through the content of the game and skills arising as a result of playing the game (McFarlane et al. 2002). As the computer games industry grows, research is being conducted on the use of these as an educational tool. However, the challenge lies in identifying the pedagogical features needed to balance the 'fun and enjoyable' aspect of a game with the appropriate content needed to raise understanding of a particular domain. Studies suggest that offering appropriate and adaptive feedback, the embedding of cognitive strategies and animated graphics which reduce task time and increase achievement, have a positive effect on student learning (Rosas et al. 2003).

Principle	*Description*	*Theoretical Rationale*
Guided Activity	Students learn better when allowed to interact with a pedagogical agent who helps guide their cognitive processing	Guided activity encourages essential and generative processing by prompting students to engage in selection, organisation and integration of new information.
Reflection	Students learn better when asked to reflect upon correct answers during the process of meaning making	Reflection promotes essential and generative processing by encouraging more active organisation and integration of new information
Feedback	Students learn better with explanatory rather than corrective feedback alone	Explanatory feedback reduces extraneous processing by providing students with proper schemas to repair their misconceptions
Pacing	Students learn better when allowed to control the pace of presentation of the instructional material	Pace control reduces representational holding by allowing students to process smaller chunks of information in working memory
Pretraining	Students learn better when they receive focused pretraining that provides or activates relevant prior knowledge	Pretraining helps guide the learners generative processing by showing which aspects of prior knowledge to integrate with incoming information

Table 2.2 Five design principles and corresponding theoretical Rationale (Moreno and Mayer, 2007)

2.5.3 Augmented reality

Pervasive technologies and Augmented Reality (AR) have provided many opportunities within educational games. Complementing the real world by proving information or enabling objects to be manipulated, AR does not alter real world experiences, but rather enhances them. Stock et al. (2005, p. 253) state, "Augmented reality systems generate a composite view for the user: a combination of the real scene perceived by the user and a virtual scene generated by the computer that augments the scene with additional information". By augmenting real objects with pervasive devices such as radio frequency identification devices (RFID) and embedded sensors, virtual information can be provided based on location and context. Hinske et al. (2007) used a Playmobil Knights' Castle toy as a platform for teaching about the Middle Ages through AR. By using a PDA and embedded RFID technology, the Augmented Knight's Castle provides verbal commentary, music and visual feedback based on the child's play. *Read-It* is a multimodal tangible interface to aid in children's reading (Weevers et al. 2004). Based upon the game 'Memory', a child turns a blank block which has an embedded sensor and the Read-It system then provides visual and auditory feedback on the spelling and pronunciation of the virtual pictographic word associated with that block. *Virtuoso* (Wagner et al. 2006) is a PDA based Art History game where players must sort a collection of artworks according to their date of creation. Based within a museum, the PDA and embodied agent provide virtual hints and information on a particular piece of artwork through audio and visual display.

2.5.4 Virtual learning environments (VLE)

A virtual learning environment is an educational tool which helps monitor students' progress online. These VLEs offer an online storage facility where students can access the information from any device connected to the internet and at any time. VLEs such as *Blackboard* (Blackboard, 2009) and *Moodle* (Moodle, 2009) enable teachers to create modules and dispense access rights to their content. Students are able to upload examples of work for correction and conduct examinations and practicals online. The use of VLEs has had a positive effect in higher education with most if not all universities utilising a VLE. As suggested by Dillenbourg (2002) the pedagogical effects of VLEs are unclear. However, they provide a means for collaboration, social interaction, distant learning and unlimited resources. Like any other tool, it is the individual's use of it which makes it effective.

2.5.5 Virtual reality

Encarta (2009, p. 1) defines Virtual Reality (VR) as “a technique by which a computer simulates a three-dimensional physical environment using visual and auditory stimuli”. Virtual reality systems provide a rich sensory experience which can be utilised in game play to immerse the player in the game environment. This can also be applied in educational systems to enhance the learning and understanding of important concepts and principles (Crumpton et al. 1997). VR systems are ideal for simulation type training which enables the simulation of high risk scenarios in a safe and controlled environment. Virtual worlds such as *SecondLife* (Linden Labs, 2009) provide a collaborative virtual environment which enables entertaining and educational uses. Residents can socialise, visit parties, seminars and even discover educational islands such as *Engineering Education Island* (Virtual worlds team UU, 2008). Engineering Education Island is a working prototype which integrates a virtual world (Second life) with a VLE *Sloodle* (Second life object oriented dynamic learning environment). The Education Island hosts three practical tasks each with a subset of mini tasks to complete. Integrating the VLE scripted sensors enables the interaction during each practical to be logged and monitored by the student and progress to be monitored by the lecturer. Figure 2.6 shows the detail collected and provided by Sloodle on activating the RC circuit, practical 2 (Callaghan et al. 2009).

2.5.6 Mobile handheld devices

Mobile handheld devices are now faster and more powerful and thus lifting the restriction of computing to a fixed physical location. Handheld devices have revolutionised the tourist industry by proving context-aware and location-aware information. These devices have also revolutionised the gaming industry and research is being investigated into their educational use (Dale et al. 2009; Lipinski, 2008). Referring to handheld devices such as the Nintendo DS and PSP, Pulman (2007, p. 535) states, “As these systems revolutionise handheld gaming, we believe their educational possibilities are also revolutionary and exciting times lay ahead for the educators and students that can harness them”. The Nintendo DS (Nintendo, 2009) and Apple’s iTouch and iPhone (Apple, 2009) appear to be leading the mobile handheld device market. The Nintendo DS (NDS) has sold over 96 million consoles to date (Faylor, 2009) and offers a dual screen, wireless connectivity and touch screen technology. Apple’s iTouch and iPhone offer multi touch technology, wireless connectivity and provide endless features available as downloads from their App Store. Many schools are currently utilising these devices to make learning easier and fun. A school in south-east London regularly uses Dr Kawashima’s Brain Training on the Nintendo DS during their mathematics lessons (Galloway, 2009).

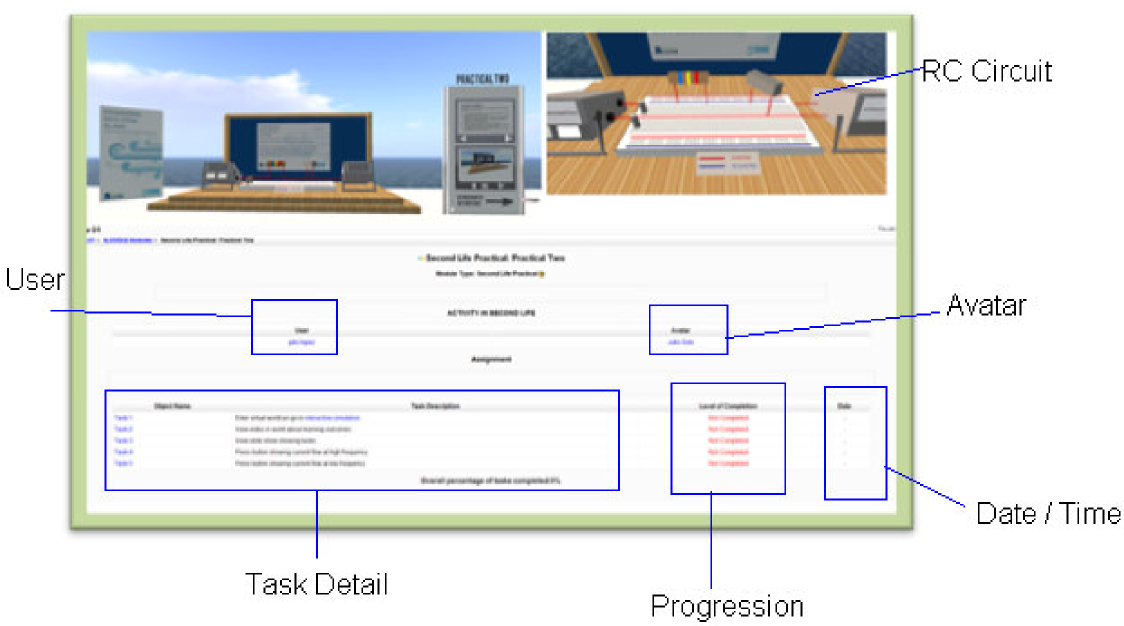

Figure 2.6 Practical 2 in Engineering Education Island integrated with Sloodle (Callaghan et al. 2009)

3 Project proposal

This section outlines the project proposal. An AmbiLearn architecture and associated client/server interaction is presented which highlights the core system activity. A TreasureLearn activity diagram outlines the user interaction with the application and testing and evaluation of AmbiLearn and TreasureLearn is given. AmbiLearn is compared to similar current multimodal learning environments and a research plan is outlined. The motivation for the project proposal has arisen from gaps in ambient multimodal learning environments in respect of children's education. Figure 3.1 provides a mind map of key concepts relating to AmbiLearn. In all key areas of related research investigations have proven successful with regards to particular educational domains. Multimodal interfaces and the use of speech and pen have proven to provide complementary interfaces supporting assistance in navigation and tourist systems. Studies have shown how embodied agents have had a positive effect on student learning by keeping students attention and hence improving motivation. Gaming devices have proven to be successful in motivating students and keeping them interested in a particular educational domain and the different learning environments have allowed a more flexible way of providing the educational content suitable to the different needs of each individual. In this research there is the potential to build on these past systems leading to the development of a multimodal game accessible via a mobile handheld device.

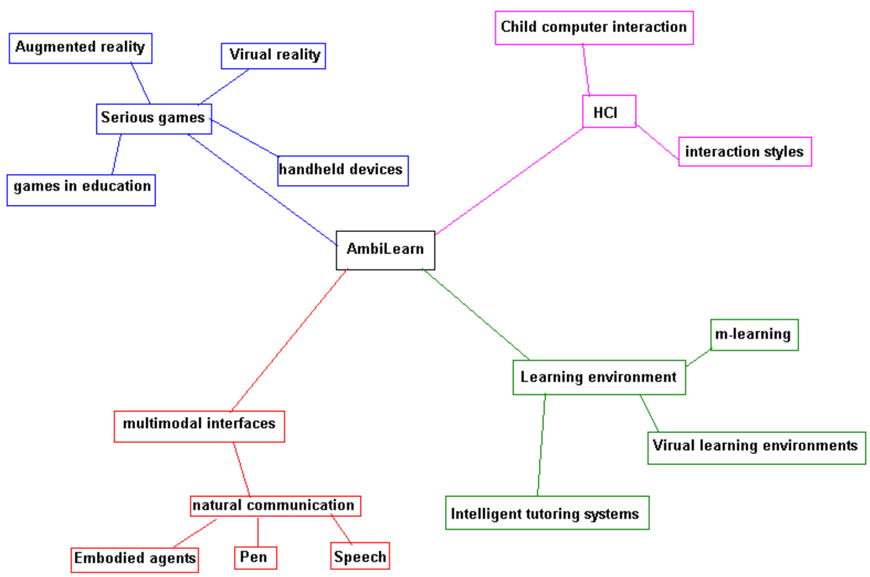

Figure 3.1 Conceptual Mind Map

A focus of this research is to further investigate the use of multimodal communication supporting collaborative learning for children using mobile handheld devices. The overall objective is the design and development of, AmbiLearn, a mobile multimodal system to facilitate a child's ubiquitous learning environment. The multimodality of AmbiLearn will enable the user to use a combination of speech and pen as input and offer rich feedback through the use of speech, graphics and an embodied agent. An application, TreasureLearn will be developed as a treasure hunt style game with a user logging component to demonstrate and test the educational potential of a multimodal, mobile game.

3.1 AmbiLearn architecture

AmbiLearn's architecture is shown in Figure 3.2. Input includes processing *speech recognition* and *gesture recognition.* A *multimodal data fusion* module fuses this information to obtain the intentions of the user. These intentions are then passed to the *dialogue manager* which decides the next action based on rules and tasks obtained from the *TreasureLearn* application. This results in output processing such as non-speech audio *generation, graphics generation* and *speech synthesis*. The dialogue manager communicates with the *multimodal data storage* module to store the multimodal data from the users' intentions and system responses. The TreasureLearn application accesses a *database* to provide pedagogical content to use in output processing, and writing to the database to store user profiles and progress.

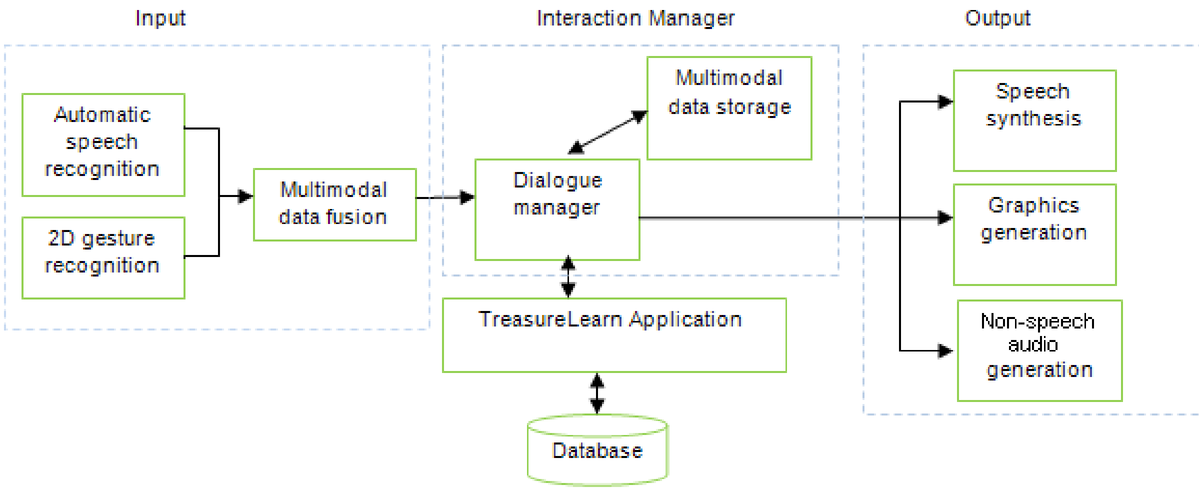

Figure 3.2 AmbiLearn Architecture

Figure 3.3 outlines the client server interaction of AmbiLearn. The input/output are split between the client and server. The function of the client side input depends on the type of mobile handheld device used. The input (client side) may be responsible for capturing and time-stamping the users' speech and 2D gesture. This may include the ASR engine for feature extraction on the incoming speech signal. The input (server side) may then handle the ASR search and integration of modalities. Implementing a client server architecture for communication purposes will support multiple users within the learning environment with each accessing the server from their own device. A peer to peer architecture among active devices will support collaborative aspects of learning, whilst deploying AmbiLearn.

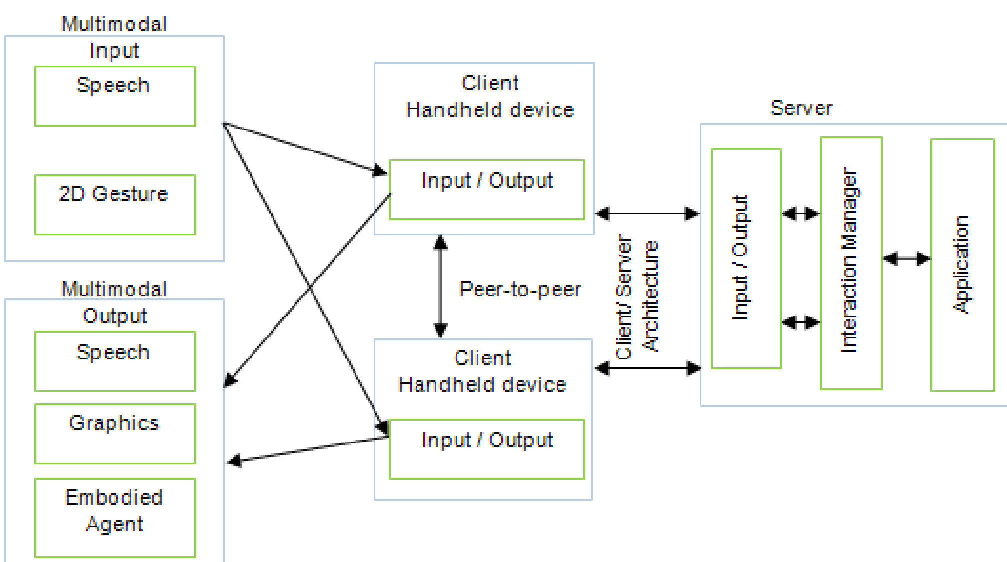

Figure 3.3 AmbiLearn Client Server Interaction

3.2 TreasureLearn application

Figure 3.4 shows a general activity diagram for TreasureLearn. The modality key outlined in Figure 3.4 provides the instances when the corresponding modality will be used. It is envisaged that navigation throughout the game will take the form of direct movement using a stylus or speech utterances or a combination of both, such as uttering 'let's go here' while pointing to the place they wish to move to. The activity modules within the TresureLearn application include providing the game play, knowledge and overall goals. The main interaction will occur in the challenge module where the users will be quizzed on their knowledge in order to pass obstacles such as locked doors or open chests. As the challenges will be mainly factual knowledge challenges, the content/actions needed to overcome the obstacles must be available within the game; therefore the pedagogical content will be provided within the knowledge module.

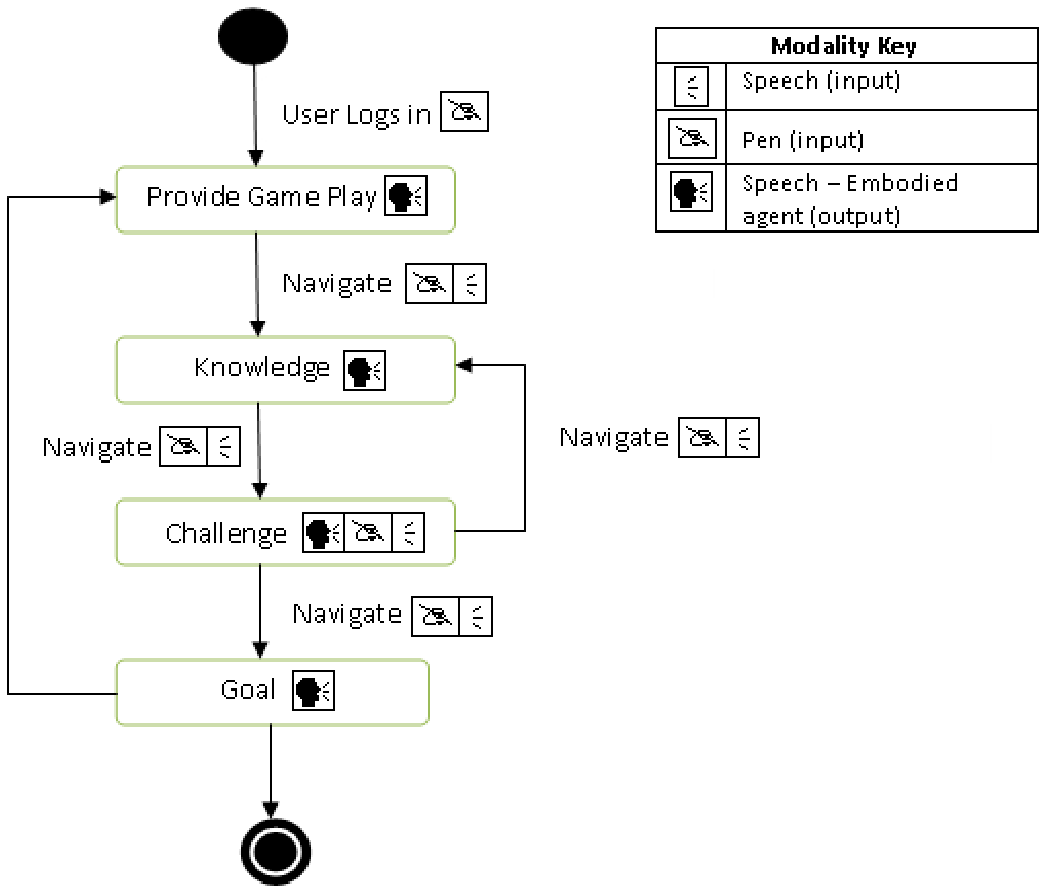

Figure 3.4 TreasureLearn Activity Diagram

Modules within TreasureLearn are as follows:

Gameplay: In game design the game play module is concerned with the challenges a player is faced with and the actions taken to overcome the challenges. This module will essentially provide the user with the rules and game concepts. This information will be provided through the use of an embodied agent, however limited interaction occurs within this module. Depending on the user and their previous interactions, this module will provide the information regarding the level, or area of game play they are due to commence.

Knowledge: A knowledge module within the game will provide the pedagogical content which will comply with the intended age group and core curriculum. This information can be provided through the embodied agent and simply provides the user with content which will need to be recalled to overcome certain challenges.

Challenge: The challenge module provides the main interaction within TreasureLearn. This module is responsible for accepting the knowledge the user provides and determining whether the user responded to an answer or puzzle correctly. The user input at this stage will be logged for analysis at a later stage.

Goal: The goal module will provide the user with an update on their progress such as moving between levels. This module is concerned with updating the user information. At the end of each game session, i.e. each time the user plays, this module will be invoked to record and update the logged information so that an external user monitoring and assessment facility is available.

3.3 Testing and evaluation

Markopoulos et al. (2008) documents a collection of methods which can be used to evaluate children's interactive products. Included are inspection methods such as SEEM (Structured Expert Evaluation Method) (Baauw et al. 2005) and survey methods, using the Fun Toolkit (Read and MacFarlane, 2006).
The Fun Toolkit, developed by Read and MacFarlane (2006) consists of three instruments which can be used by children to gather opinions on products. The Smileyometer illustrated in Figure 3.5 (a) is a visual analogue scale based around a 1 – 5 Likert scale. This tool can be used before and after interaction with any system by gathering children's expectations and their opinions. The Fun Sorter, illustrated in Figure 3.5 (b) compares a set of related technologies or features whilst the Again Again table, Figure 3.5 (c) identifies whether the child would like to use the product or feature again. To identify an educational potential of AmbiLearn, pre-test and post-test scores may be gathered. By using two groups with only one using AmbiLearn, it may be possible to identify whether using the system, helped the student retain knowledge.

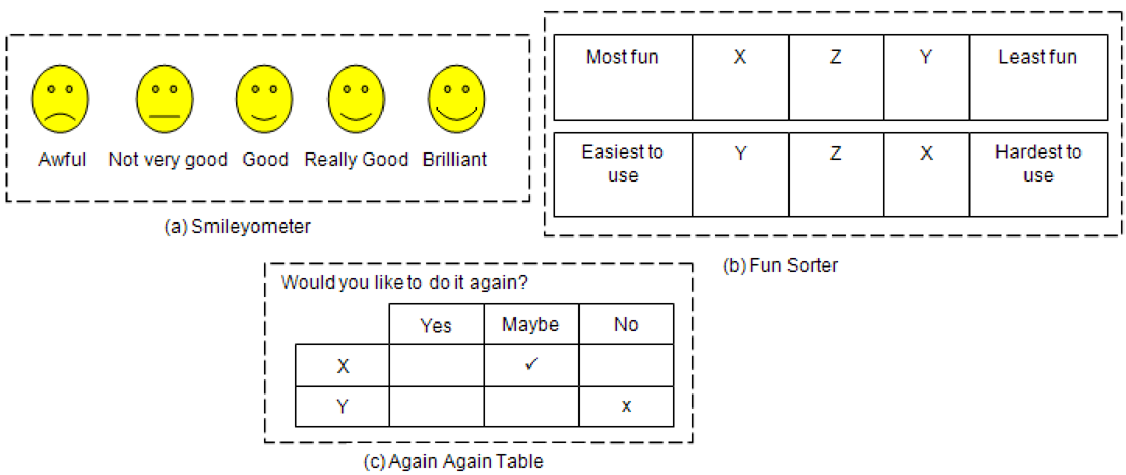

Figure 3.5 Tools used within the Fun Toolkit (Markopoulos et al. 2008)

3.4 Relation to other work

Table A.1, Appendix A includes a comparison of the proposed system AmbiLearn with existing multimodal learning environments. Table A.1 provides features of existing systems and highlights with the use of (✓) or (×) for features which the corresponding system has, or does not have, respectively. Table A.1 shows that the use of Speech as an input modality is widely available in desktop applications with most supported by an embodied agent. The development of AmbiLearn aims to follow this trend and since it is based on a mobile handheld device it is envisioned that AmbiLearn will be extended to mobile environments. AmbiLearn improves upon these existing systems by supporting collaborative learning indirectly thought the use of a game and providing a user logging module for user analysis. With an emphasis on multimodal input and output it is expected that AmbiLearn will advance the area in terms of providing a flexible and rich learning experience.

3.5 Research Plan

Appendix B, Table B.1 shows a Gantt chart outlining the plan of proposed work. A poster will be presented at the research graduate school dinner in March/April, 2010 and a 3rd year presentation in January, 2011 at the graduate school research conference. A questionnaire will be distributed in early September in order to conduct requirement analysis regarding existing use of games, gaming devices and learning environments amongst children at a primary level. Throughout the project, background literature will be investigated and activities involving the development and testing of AmbiLearn completed. The final thesis will be completed and submitted in October, 2011. Appendix C provides a preliminary outline of the final thesis. Appropriate conferences and journals will be targeted for publication over the research period.

4 Prospective tools and software analysis

Prospective software tools are being investigated with regard to the development of AmbiLearn. These include the use of toolkits and platforms supporting spoken dialogue systems, multimodal systems, embodied agents and fusion of the different modalities.

4.1 Tools for multimodal interfaces

Microsoft's .Net Speech SDK (Microsoft, 2008) offers developers APIs and tools to build their own speech-enabled applications using the Speech Server Developer tools for Visual Studio 2005. WebSphere Voice Toolkit (IBM, 2009) offers a development environment with a voice XML editor and debugging tools and a communication flow builder amongst others. XFace (Balci, 2004) is a MPEG-4 based open source toolkit for 3D Facial Animation. This toolkit offers a platform independent suite of components including an XFace library, XFace Editor and an XFace Player. COLLAGEN (for COLLaborative AGENt) is Java Middleware for building collaborative interface agents (Cassell et al. 2001). Developed at the MIT Media Lab, COLLAGEN aids in the development of agents which both communicate with and observe the actions of the user (Rich et al. 1998). Psyclone is a platform which bridges platforms and programming languages. Based on modularity principles inherited from the CDM (Constructivist Design Methodology), Psyclone simplifies the design of and implementation of systems with many components that interact in complex ways. Psyclone incorporates OpenAIR, which is an information and exchange and routing specification. OpenAIR allows flexible component interaction as it provides a language-independent messaging format and a network-independent routing protocol (Thórisson, 2007; Mindmakers, 2007).

4.2 Mobile handheld device development kits

There are many available toolkits to aid in developing on mobile handheld devices which include J2ME (Sun Developer Network, 2009). The Java Platform Micro Edition (Java ME aka J2ME) is specifically designed for development for handheld devices such as mobile phones and PDA's. The platform provides configuration options designed to run on devices with limited processing power, memory and lower graphics. The Java ME SDK is a free download complete with emulators and libraries for all mobile development needs. The homebrew communities have reverse engineered the Nintendo DS providing an open source development kit (DevKit Pro, 2009) with a plug-in which allows the source code to be scripted with the Visual Studio suite. DevKit Pro is a collection of development tool chains for a variety of gaming systems. Included in this collection is devkitARM which allows for the compiling of ARM binaries specific for development on the Nintendo DS. This development kit is complete with libraries such as libnds, libfat and PAlib which allow for access to specific NDS constructs and functions (Schroeder and Walker, 2008). NDS emulators are available which allow execution on the PC speeding up development time. NoPass devices or R4DS chips allow the running of a developed game on the Nintendo DS device. These devices include a micro SD card which allows the passing of files from PC to the device which can then be easily slotted into the game cartridge slot on the Nintendo DS. Apple (iPhone Developer Centre, 2009) have release a developers SDK in order to aid development for the iPod, iTouch and iPhone. The iPhone developer program proves Xcode, a complete development environment with source editor and graphical debugger, an Interface builder allowing easy interface design and iPhone simulator and extra instruments to collect, display and compare data performance graphically. Applications can then be transferred to the device for testing by simply plugging in the iPhone or iPod.

5 Conclusion

This report provides an overview of this research project involving the design and implementation of a multimodal interface for children deployed on a mobile handheld device. The objectives of the research include investigating methods of merging multiple modalities on a mobile handheld device. Address the interface design and usability of AmbiLearn and TreasureLearn. Consider the role of serious games in children's education and demonstrate and test the educational potential of AmbiLearn and TreasureLearn. A literature review of the key concepts has been provided within the fields of Multimodal Interfaces and Learning Environments. The project proposal is provided as well as the AmbiLearn architecture and TreasureLearn application. AmbiLearn is related to other work and aims to improve upon the existing systems by supporting collaborative learning indirectly thought the use of a game and providing a user logging module for analysis. With an emphasis on multimodal input and output it is expected that AmbiLearn will advance the area in terms of providing a flexible and rich learning experience for children. A review of potential tools has highlighted the use of toolkits and platforms supporting spoken dialogue systems, multimodal systems, embodied agents and fusion of the different modalities.

Future work includes investigating further which mobile handheld devices are being used by children and whether these devices are being used to support learning. From results gathered an insight will be gained into what games children are using and what features make them appealing. Knowledge of what learning environments offer most potential at a primary school level and how the children feel about using a system such as AmbiLearn will be gained. The information obtained will provide an insight into the user requirements of AmbiLearn and the TreasureLearn game and design the system to meet them. Further investigation into the prospective tools available to aid in the development of AmbiLearn is needed as well as designing the infrastructure with regard to integration, decision making and platform characteristics. Continuous investigations of literature within the key fields of multimodal interfaces and learning environments will be done throughout the duration of the project and I plan to target appropriate conferences and journals to publish the results.

Appendix A: Comparison of multimodal educational systems

System	Year	Device		Input Modality			Output modality			Embodied Agent	Support collaboration	User Logging	Educational Domain
		Desktop PC	Mobile	Speech	Touch	visual	Speech	Graphics	Text				
Gandalf	1997	✓	✗	✓	✗	✓	✓	✓	✓	✓	✓	✗	Solar system
Steve	1997	✓	✗	✓	✓	✗	✓	✓	✗	✓	✓	✓	Virtual Procedural Training
KidsRoom	1998	✓	✗	✓	✗	✓	✓	✓	✓	✗	✓	✗	Imaginative Story Telling
Storyteller	1999	✓	✗	✓	✗	✓	✓	✓	✗	✗	✗	✗	Story Telling
TICLE	2002	✓	✗	✗	✓	✓	✓	✓	✗	✗	✓	✗	Tangram Puzzle
Read-It	2004	✓	✗	✗	✗	✓	✓	✓	✓	✗	✓	✗	Reading and Language
NICE	2005	✓	✗	✓	✓	✗	✓	✓	✗	✓	✗	✗	Life and Works of Hans Christian Anderson
Virtuoso	2006	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓	✗	Art History
Augmented Knights Castle	2007	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓	✗	History (Middle Ages)
AmbiLearn	2011	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	Treasure Hunt / Puzzles

Table A.1 Comparison of multimodal educational systems

Appendix B: Proposed research schedule

	Activities
	Submissions
	Deliverables

Research Activities	**2008**	**2009**				**2010**				**2011**		
	Oct-Dec	**Jan-Mar**	**Apr-Jun**	**Jul-Sep**	**Oct-Dec**	**Jan-Mar**	**Apr-Jun**	**Jul-Sep**	**Oct-Dec**	**Jan-Mar**	**Apr-Jun**	**Jul-Sep**
Perform literature review												
100 Day Review and VIVA												
Develop AmbiLearn & TreasureLearn (design/implementation)												
Confirmation report												
PGNET 2009 Conference												
Initial Questionnaire												
2^{nd} Year Poster												
Test and Evaluate AmbiLearn												
3^{rd} Year Presentation												
Thesis write up												
Submit Thesis												

Table B.1 Proposed research schedule

Appendix C: Preliminary thesis structure

Abstract
Acknowledgements
Declaration

1 Introduction

2 Literature Review

3 Ambient Intelligent Multimodal Learning Environment for Children

4 AmbiLearn and TreasureLearn Implementation

5 Evaluation of AmbiLearn and TreasureLearn

6 Conclusion

7 Appendices

8 References

References

Anegg, H., Niklfeld, G., Pucher, M., Schatz, R., Simon, R., Wegscheider, F., Dangl, T and Jank, M. (2004) Multimodal Interfaces in Mobile Devices – The Mona Project. In: *Proceedings of the workshop on emerging applications for wireless and mobile access, 13th International World Wide Web Conference NY*.

Apple (2009) Available from: http://www.apple.com/ipodtouch/

Balci, K. (2004) Xface: MPEG-4 Based open Source Toolkit for 3D Facia Animation. *Working Conference on Advanced Visual interfaces, Gallipoli, Italy, May.* New York: ACM, 399-402.

Baillie, L., Schatz, R., Simon, R., Anegg, H., Wegscheider, F., Niklfeld, G. and Gassner, A.(2005) Designing Mona: User Interactions with Multimodal Mobile Applications. In: *Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas, Nevada, USA, July.*

Bernsen, N.O., H. Dybkjær and L. Dybkjær (1998) *Designing Interactive Speech Systems. From First Ideas to User Testing.* London: Springer.

Blackboard (2009) Available from: http://www.blackboard.com/

Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, Y., Schütte, A. and Wilson, A. (1998a) Perceptual user interfaces: The KidsRoom. *Communications of the ACM*, 43(3), 60-61.

Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, L., Ivanov, Y., Schütte, A. and Wilson, A. (1998b) *The kidsroom: A perceptually-based interactive and immersive story environment.* Technical report, MIT Media Laboratory, The MIT Media Laboratory, 20 Ames Street, Cambridge, MA 02139.

Bobick, A., Intille, S., Davis, J., Baird, F., Pinhanez, C., Campbell, Y., Schütte, A. and Wilson, A. (2000) Perceptual user interfaces: The KidsRoom. *Communications of the ACM*, 43(3), 60-61.

Bolt, R. A. (1980) Put-that-there: Voice and gesture at the graphics interface. In *Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, Seattle, Washington, United States, July.* New York: ACM, 262-270.

Booth, R. (2009) Video games are good for children – EU report. The Guardian, Thursday 12th Feburary. 2.

Brøndsted, T., P. Dalsgaard, L.B. Larsen, M. Manthey, P. Mc Kevitt, T.B. Moeslund and K.G. Olesen (2001) The IntelliMedia Workbench – an environment for building multimodal systems, In *Cooperative Multimodal Communication*. LNCS 2155, Berlin/Heidelberg: Springer, 217-233.

Bruckman, A., A. Bandlow and A. Forte (2007) HCI for Kids. In: A. Sears and J. Jacko, (eds.) *The Human-Computer Interaction Handbook. Fundamentals, evolving technologies and emerging applications.* Second edition. New York, USA: CRC press Inc. 793-810.

Buisine, S. and Martin, J. (2005) Children's and Adult's Multimodal Interaction with 2D Conversational Agents. In: *CHI'05 Extended Abstracts on Human Factors in Computing System, Portland, Oregon, USA, April.* New York: ACM, 1240-1243.

Callaghan, M. and McCusker, K. (2009) *Worlds in Motion, Rise of Virtual Worlds*. Unpublished.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H. and Yan, H. (1999) Embodiment in conversational interfaces: REA. *SIGCHI Conference on Human Factors in Computing Systems: the CHI is the Limit, Pittsburgh, Pennsylvania, United States, May.* New York: ACM, 520-527.

Cassell, J., Nakano, Y.I., Bickmore, T., Sidner, C.L. and Rich, C. (2001) Non-Verbal Cues for Discourse Structure. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, Toulouse, France, July.* NJ: ACL, 114-123.

Charlier, J., Buquet, C., Dubus, F., Hugeux, J. and Degroc, B. (1997) VISIOBOARD: a new gaze command system for handicapped subjects. *Medical and Biological Engineering and Computing.* 35, 416 supplement.

ChiCI Group (2007) *Child Computer Interaction* [online]. Available from: http://www.chici.org/about_us.php [Accessed 04/11/08].

Cohen, P. (1992) The Role of Natural Language in a Multimodal Interface. In: *Proceedings of the 5th Annual Symposium on user interface software and technology, Monterau, California, US.* New York: ACM, 143–149.

Cohen, P., Johnston, M., McGee, D., Oviatt, S., Pittman, J., Smith, I., Chen, L., and Clow, J. (1997) QuickSet: multimodal interaction for distributed applications. In *Proceedings of the Fifth ACM international Conference on Multimedia , Seattle, Washington, United States, November 09 – 13.* New York: ACM, 31-40.

Corradini, A., Mehta, M., Bersen, N. And Charfuelan, M. (2005) Animating an interactive conversational character for an educational game system. In *Proceedings of the 10th international Conference on Intelligent User Interfaces, California, January.* New York: ACM, 183-190.

Crumpton, L. and Harden, E. (1997) Using Virtual Reality as a tool to Enhance Classroom Instruction. In: *Proceedings of the 21st International Conference on computers and Industrial Engineering*. 33(1-2). 217-220.

Dale, C. and Pymm, J. (2009) Podagogy, The iPod as a learning technology. In: *Active Learning in Higher Education*. Sage, 10(1). 84-96.

Delgado, R.L.C. and M. Araki (2005) *Spoken, Multilingual and Multimodal Dialogue Systems, Development and Assessment.* Chichester, West Sussex, England: Wiley.

DevKit Pro (2009) Available from: http://www.devkitpro.org/downloads/

Dillenbourg, P., Schneider, D. and Synteta, P. (2002) Virtual Learning Environments. In: *Proceedings of the 3rd Hellenic Conference "Information & Communication Technologies in Education".* Greece: Kastaniotis. 3-18.

Dusan, S., Gadbois, G.J. and Flanagan, J. (2003) Multimodal Interaction on PDA's Integrating Speech and Pen Inputs. In: *Eurospeech 2003, 8th European conference on Speech Communication and Technology, Geneva, Switzerland, September*. ISCA, 2225–2228.

Encarta (2008) Available from: http://encarta.msn.com/dictionary_/virtual%2520reality.html

Faylor, C. (2009) Latest Worldwide Console Sales Revealed [online]. Available from: http://www.shacknews.com/onearticle.x/56993 [Accessed 06/04/09].

Galloway, J. (2009) *Harnessing technology for every child matters & personal learning.* Oxon: Routledge.

Hinske, S. and Lampe, M. (2007) Interactive Educational Play with Augmented Toy environments. *ECRIM News,* 71, 38-40.

Hutchinson, T., White, K., Martin, W., Reichert, K. and Frey, L. (1989) Human computer interaction using eye gaze input. *IEEE Transactions on Systems, Man and Cybernetics, Nov/Dec,* 19(6), 1527-1534.

Hyndman, J., Lunney, T. and McKevitt, P. (2009) AmbiLearn: Ambient Intelligent Multimodal Learning Environment for Children. In: *Proceedings of the 10th Annual Convergence of Telecommunications, Networking & Broadcasting, PG Net, Liverpool, June 22nd – 23rd*.

IBM (2009) Available from: http://www-01.ibm.com/software/pervasive/voice_toolkit/

iPhone Developer Centre (2009) Available from: http://developer.apple.com/iphone/

Johnson, W.L. and Rickel, J. (1997) Steve: an animated pedagogical agent for procedural training in virtual environments. *SIGART Bulletin, December,* 8 (1-4), 16-21.

Koons, D., Sparrell, C. and Thorisson, K. (1993) Integrating simultaneous input from speech, gaze, and hand gestures. In: M. Maybury, Ed. *Intelligent Multimedia interfaces*, M. Maybury, Ed. Menlo Park, CA: American Association for Artificial Intelligence. 257-276.

Lester, J.C., Converse, S.A., Kahler, S.E., Todd Barlow, S., Stone, B.A. and Bhogal, R.S. (1997) The Persona Effect: Affective Impact of Animated Pedagogical Agents. In *Proceedings of the SIGCHI conference on Human factors in computing systems, Atlanta, Georgia, United States, March.* New York: ACM, 359-366.

Linden Lab (2009) Available from: http://secondlife.com/

Lipinski, M. (2008) *Feasibility of the Nintendo DS for teaching problem-based learning in kindergarten through twelfth grade students.* MSc Thesis, University of Missouri-Columbia.

Malaka, R., Haeussler, J. and Aras, H. (2004) SmartKom mobile: intelligent ubiquitous user interaction. In: *Proceedings of the 9th international Conference on intelligent User interfaces, Madeira, Portugal, January 13 - 16.* New York: ACM. 310-312.

Markopoulos, P., Read, J., MacFarlane, S. and Höysniemi, J. (2008) *Evaluating children's interactive products. Principles and Practices for Interaction Designers.* Burlington, USA: Morgan Kaufmann.

Martin, J., Buisine, S., Pitel, G., Bernsen, N.O. (2006) Fusion of children's speech and 2D gestures when conversing with 3D characters. *Multimodal Human-computer Interfaces*, *Signal Processing* 86 (12). 3596-3624.

Maybury, M. (1993) Intelligent Multimedia Interfaces. California, USA: AAAI Press.

Maybury, M. (1998) Intelligent user interfaces: an introduction. In *Proceedings of the 4th international conference on Intelligent user interfaces, Los Angeles, California, United States, January*. New York: ACM, 3-4.

McFarlane, A., Sparrowhawk, A. And Heald, Y. (2002) *Report on the educational use of games: An exploration by TEEM of the contribution which games can make to the education process* [online]. Available from: http://www.teem.org.uk/publications/teem_gamesined_full.pdf [Accessed 03/11/08].

Mc Kevitt, P. (Ed.), (1995-96) *Integration of Natural Language and Vision Processing, (Vols. I – IV),* Dordrecht, The Netherlands: Kluwer Academic Publishers.

Mc Tear, M. (2004) *Spoken Dialogue Technology. Towards the conversational user interface*. London, England: Springer.

Microsoft (2008) Available from: http://www.microsoft.com/speech/speech2007/default.mspx

Mindmakers (2007) *Psyclone* [online]. Available from: http://www.mindmakers.org/projects/Psyclone [Accessed 15/11/08].

Moodle (2009) Available from: http://moodle.org/

Moreno, R. and Mayer, R. (2007) Interactive Multimodal Learning Environments. *Educational Psychology Review, Special issue on Interactive Learning Environments: Contemporary Issues and Trends*. Springer 19. 309-326.

Niklfeld, G., Anegg, H. and Gassner, A. (2005) Device independent mobile multimodal user interfaces with the MONA multimodal Presentation Server. In: *Proceedings of Ubiquitous Services and Applications Exploiting the Potential, Eurescom Summit , Heidelberg, Germany, April.*

Nintendo (2009) Available from: http://www.nintendo.co.uk/NOE/en_GB/nintendo_ds_family_page_11512.html

O'Malley, C., Vavoula, G., Glew, J., Taylor, J., Sharples, M. and Lefrere, P. (2003) *MOBILearn WP4 – guidelines for Learning/Teaching/Tutoring in a Mobile Environment* [Online]. MobiLearn Project. Available from: http://www.mobilearn.org/download/results/guidelines.pdf [Accessed 03/04/09].

Oviatt, S. (1999) Ten Myths of Multimodal Interaction. *Communications of the ACM, November*. New York: ACM. 42 (11). 74-81.

Oviatt, S. (2007) Multimodal Interfaces. In: A. Sears and J. Jacko, (eds.) *The Human-Computer Interaction Handbook. Fundamentals, evolving technologies and emerging applications.* Second edition. New York, USA: CRC press Inc, 413-432.

Oviatt, S., Cohen, P., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J. and Ferro, D. (2000) Designing the user interface for multimodal speech and pen-based gesture applications: state-of-the-art systems and future research directions. *Human Computer Interaction.* NJ, USA: L. Erlbaum Associates. 15(4). 263-322.

Ortiz, A., Carretero, M., Oyarzun, D., Yanguas, J.J., Buiza, C., Gonzalez, M.F. and Etxeberria, I. (2007) Elderly Users in Ambient Intelligence: Does an Avatar Improve the Interaction?. *Universal Access in Ambient Intelligence Environments.* LNCS 4397, Berlin/Heidelberg: Springer, 99-114.

Prendinger, H., Chunling, M., Yingzi, J., Nakasone, A. and Ishizuka, M. (2005) Understanding the Effect of Like-like Interface Agents Through Users' Eye Movements. In: *Proceedings of the 7th international conference on Multimodal interfaces. Torento, Italy October,* New York: ACM. 108-115.

Pullman, A. (2007) Can a handheld gaming device be used as an effective assistive technology tool? *British Journal of Educational Technology, Oxford, UK: Blackwell.* 38(3). 534-535.

Read, J. and MacFarlane, S. (2006) Using the Fun Toolkit and Other Survey Methods to Gather Opinions in Child Computer Interaction. In: *Proceedings of the 2006 conference on interaction design and children, Tampere, Finland, June.* New York: ACM, 81-88.

Rich, C., Lesh, N., Rickel, J. And Garland, A. (1998) A Plug-in Architecture for Generating Collaborative Agent Responses. In *Proceedings of the First international Joint Conference on Autonomous Agents and Multiagent Systems, Italy, July.* New York: ACM, 782-789.

Rickel, J. and Johnson, W. (1999) Animated agents for procedural training in virtual reality: Perception, Cognition and Motor Control. *Applied Artificial Intelligence.* Taylor & Francis. 13, 343-382.

Rosas, R., Nussbaum, M., Cumsille, P., Marianov, V., Correa, M., Flores, P., Grau, V., Lagos, F., Lopez, X., López, V., Rodriguez, P. and Salinas, M. (2003) Beyond Nintendo: design and assessment of educational video games for first and second grade students. *Computers & Education,* 40(1), 71-94.

Santamaria, M. (2003) Web Navigation recorder eyetracking software application demo In: *Proceedings of the 2003 conference on Interaction design and children, Preston, England*, NY: ACM. 166.

Scarlatos, L.L. (2002) TICLE: using multimedia multimodal guidance to enhance learning. *Information Sciences,* 140 (1-2), 85-103.

Schroeder, C. and Walker, L. (2008) *Homebrew Computing and the Nintendo DS: Programming and Controlling a Dedicated-Purpose Computer system*. [online] CIT 595, University of Pennsylvania. Available from: http://www.cis.upenn.edu/~palsetia/cit595s08/Projects08/DSHomebrewComputing.pdf [Accessed 12/01/09].

Scott, S.D., Mandryk, R.L. and Inkpen, K.M. (2003) Understanding children's collaborative interactions in shared environments. *Journal of Computer Assisted Learning,* 19(2), 220-228.

Shadbolt, N. (2003) Ambient Intelligence. *IEEE Intelligent Systems,* 18(4), 2-3.

Stock, O. and Zancanaro, M. (2005) *Multimodal Intelligent Information Presentation*. Dordrecht, the Netherlands: Springer.

Strommen, E., Revelle, G., Medoff, L. and Razavi, S. (1996) Slow and steady wins the race? Three-year old children and pointing device use. *Behaviour & Information Technology*. Taylor & Francis. 15(1). 57-64.

Sun Developer Network (2009) Available from: http://java.sun.com/javame/index.jsp

Thórisson, K.R. (2007) Integrated A.I. systems. *Minds and Machine,* 17(1), 11-25.

Thórisson, K. (2005) *Gandalf: The communicative Humanoid* [online] http://alumni.media.mit.edu/~kris/gandalf.html#what-is [Accessed 04/03/09].

Thórisson, K.R. (1997) Gandalf: an embodied humanoid capable of real-time multimodal dialogue with people. *First international Conference on Autonomous Agents, Marina del Rey, California, United States, February.* New York: ACM, 536-537.

Thórisson, K. (1996) *Communicative Humanoids: A Computational Model of Psychosocial Dialogue Skills* [online]. Ph.D. Thesis, Massachusetts Institute of Technology. Available from: http://xenia.media.mit.edu/~kris/thesis.html [Accessed 04/03/09].

USC (2000) *Steve in Action* [online] University of Southern California. Available at: http://www.isi.edu/isd/VET/steve-demo.html [Accessed: 9/03/09].

Virtual worlds team UU. (2008) Available from: http://scis.ulster.ac.uk/Virtual-Worlds-Research.html

Wagner, D., Schmalstieg, D. and Billinghurst, M. (2006) Handheld AR for Collaborative Edutainment. In *Advances in Artificial Reality and Tele-Existence.* LNCS 4282, Berlin/Heidelberg: Springer, 85-96.

Wahlster, W. (2006) Dialogue Systems go multimodal: The SmartKom experience. In: W. WAHLSTER, (ed.) *SmartKom: Foundations of Multimodal dialogue Systems.* Berlin/Heidelberg: Springer-Verlag, 3-28.

Wahlster, W. Reithinger, N. and Blocher, A. (2001) SmartKom: Towards multimodal dialogues with anthropomorphic interface agents. In: *Proceedings of International Status Conference "Human-Computer Interaction", Berlin, Germany, October.* 23-34.

Weevers, I., Sluis, W., Van Schijndel, C., Fitrianie, S., Kolos-Mazuryk, L. and Martens, J.B. (2004) Read-It: A Multi-modal Tangible Interface for Children Who Learn to Read. *Entertainment Computing – ICEC.* LNCS 3166, Berlin/Heidelberg: Springer, 226-234.

Witkowski, M., Arafa, Y. and de Bruijn, O. (2001) Evaluating User Reaction to Character Agent Mediated Displays using Eye-tracking Technology. In: *Proceedings AISB-01 Symposium on Information Agents for Electronic Commerce.* 79 – 87.

Zaykovskiy, D. (2006) Survey of the speech recognition techniques on mobile devices. In: *Proceedings of SPECOM, St Petersburg, June.* 88-93.